

LIFE AND ENERGY

Isaac Asimov



AVON PUBLISHERS OF BARD, CAMELOT, DISCUS AND EQUINOX BOOKS

TO RICHARD K. WINSLOW and TIMOTHY SELDES,
gentlemen,
scholars,
and bon vivants.

AVON BOOKS
A division of
The Hearst Corporation
959 Eighth Avenue
New York, New York 10019

Copyright © 1962 by Isaac Asimov.
Published by arrangement with Doubleday & Company, Inc.
Library of Congress Catalog Card Number: 61-12491.

All rights reserved, which includes the right
to reproduce this book or portions thereof in
any form whatsoever. For information address
Doubleday & Company, Inc., 277 Park Avenue,
New York, New York 10017

First Printing (Discus Edition), January, 1972

DISCUS BOOKS TRADEMARK REG. U.S. PAT. OFF. AND
FOREIGN COUNTRIES, REGISTERED TRADEMARK—
HECHO EN CHICAGO, U.S.A.

Printed in the U.S.A.

Contents

PART I: ENERGY

1	<i>The Effort We Make</i>	1
2	<i>The Consequences of Fire</i>	13
3	<i>The Measurement of Motion</i>	26
4	<i>The Measurement of Heat</i>	39
5	<i>The Flow of Heat</i>	49
6	<i>The Moving Particles</i>	61
7	<i>The Particles Hold Together</i>	76
8	<i>The Heat of Reaction</i>	91
9	<i>The Direction of Reaction</i>	105
10	<i>The Electrical Route</i>	118
11	<i>The Encouragement of Reaction</i>	132
12	<i>Encouragement without Energy</i>	151

PART II: THE BODY

13	<i>Life and Non-Life Again</i>	162
14	<i>The Idling Speed</i>	175
15	<i>The Catalysts of Life</i>	189
16	<i>The Detail of Protein</i>	199
17	<i>The Weak Attraction</i>	218
18	<i>The Workings of Enzymes</i>	237
19	<i>The Outer Edges</i>	257
20	<i>Life without Air</i>	273
21	<i>The Importance of Phosphate</i>	282
22	<i>The Transfer of Electrons</i>	298
23	<i>Life with Air</i>	311
24	<i>The Joining of the Routes</i>	331
25	<i>The Ultimate Source</i>	350
	<i>Index</i>	367

Chapter I

THE EFFORT WE MAKE

Man is very aware, naturally, of the difference between himself and other objects in the world about him. In almost all cultures, he considers himself the peak of creation; less, perhaps, than gods, demons, angels, and other beings that are more than natural; but certainly greater and more important than anything he can see and touch.

Yet the difference between man and non-man is not an unvarying gulf. There exist degrees of separation. There are many things in great variety that share with man the faculty we call "life," which are by that very fact less removed from ourselves than the many other things that do not possess life.

Life and non-life may seem to us (from our prejudiced point of view) to be surely the most significant way in which we can divide the universe into two parts. And if we think about it superficially (the more superficially, the better) there is no difficulty at all in distinguishing between living things and non-living things. A rock in the field is certainly not alive. A lizard scurrying over the rock is certainly alive.

How can we tell? Well, the rock is lying there inert, immobile. It may be subjected to the battering of its environment, but it does not respond. The lizard, on the other hand, is moving rapidly. It is responding to some factor in its environment: to the sun, to hunger, to danger; by searching for shade or for food or for a hiding place.

Granted there is no problem, then, in distinguishing lizards (or flies or sparrows or woodchucks) from rocks, by virtue of the movement and activity of the living creatures. Are

we nevertheless certain we can tell an oyster from a rock, or that we can tell a small seed from a grain of sand equal in size and color? Can we feel confident that the first, in either case, is alive and the second not?

Actually, we can. An oyster may seem quiescent to us, but it can open its shell, draw water into its gut, strain out food particles, eliminate the residue. Compared to a rock, it is a miracle of activity. As for a seed, although it apparently does nothing at all, it needs only the proper environment—soil, water, an appropriate temperature—to become active. It will send out shoots and roots and become a plant. Compared to the grain of sand, the seed, too, is a miracle of activity.

But to do something in response to environmental factors, rather than simply to suffer its blows passively, requires the making of an effort. This "making of an effort" is a rather subtle concept I shall come back to later in the book, but for the moment let us accept the phrase at face value and in its most familiar meaning.

For instance, there is no doubt that when we run, or climb, or lift a weight we are making an effort. We are also constantly making an effort if we simply lie quietly in bed, inasmuch as we lift the weight of our rib cage with every breath we take, push a weight of blood through our body with every beat of our heart. Our kidneys and liver, and other organs as well, engage in activities of which we are not directly conscious but which involve making an effort.

Even the quiescent plant world makes what we would instantly accept as an effort: when a flower opens, when a shoot extends upward toward the light or a root makes its way downward toward water.

No rock, no non-living object, makes an effort in the same sense that even the simplest form of life does. At least our common experience would tell us so and we might therefore begin by assuming that it is the ability to make an effort that distinguishes life from non-life. Further, we might assume that the loss of the ability is what we mean by death.

In order to understand the nature of life, then, and, in particular, the workings of the human body as a specific example of life, it would seem wise to explore just what

is meant and implied by the phrase "making an effort." It is to that exploration that this book is devoted.

If we begin by assuming that all living objects can make an effort, we can next ask in what way man can differentiate himself, in this respect, from other living creatures.

The distinction that might first spring to mind is that mankind, being more intelligent than other living creatures, can organize his effort with a clearer perception of future contingencies than can any other species. Man will plant grain and tend it assiduously and with great effort for many months, although there is no immediate return at all. Eventually, however, this effort will assure him a food supply during the famine time of winter when other animals must migrate, hibernate, make do on reduced rations, or starve.

Mankind has used numerous words to describe this purposeful effort. In English we have words like "work," "toil," "labor," "drudgery," "striving," and so on. The word used by physicists to describe the usual result of making an effort is the most common of these: "work."

This, in a sense, is unfortunate, for in ordinary language there is a strong psychological component to the significance of "work." More of an effort may be involved in playing a game of tennis than in alphabetizing a set of file cards, but we consider the latter "work" and the former anything from "fun" to "competing" to "exercise" but never "work." Again, we may strive to lift a heavy weight and fail and declare that this was "hard work."

Scientists cannot allow psychological interpretations to alter the terms they use, if they are to avoid confusion. In order to make "work" meaningful, they have had to give it a mathematical definition that is independent of human opinion. This definition requires the scientist to say that, when you strive and fail to lift a heavy weight, you have done no work on that weight.

Naturally the difference between work in the scientific sense and work in the ordinary sense may puzzle the beginner and it would have been better if the scientist had made up a new word for his purpose. However, this was not done and that's that. Eventually I shall try to explain the scientific meaning of work, but for the moment we can

get along with our intuitive understanding that work is that which is usually the result of making an effort.

Scientists have also found it convenient to make use of a term to describe the capacity of any object to do work. For this purpose the word "energy" was adopted. Originally it meant simply "activity" and comes from a Greek word *energos* meaning "active." However, the Greek word breaks down to *en* (in) and *ergon* (work), so that something with energy can be thought of as having "work in" it. Again, the scientific definition of energy is precise and mathematical, but again we can rely, for the moment, on our intuitive understanding of the term.

Energy is something that is required if a living thing is to be capable of "making an effort" but it is of wider significance too. A non-living thing may not be capable of making an effort but it can contain energy. A rock teetering at the edge of a cliff cannot make the effort to fall off, but if it is pushed by a human hand (or by the wind) it can demolish a building at the foot of the cliff when it strikes. It has performed work and therefore it had contained energy.

In this sense the lizard and the rock (with which I opened the book) differ in that the energy possessed by the lizard can be channeled into work at its own will to suit its own need, whereas any energy possessed by the rock cannot be channeled into work except in response to some external force. This is a more roundabout, but better, way of making the distinction between "making an effort" and "not making an effort." We shall arrive at still better ways of making the distinction before we are done with the subject.

But now, with the addition of "work" and "energy" to our vocabulary, let us return to our distinction between man and other forms of life. Man expends energy and performs work in order to provide for the future, we can now say.

But is this truly a distinction between man and other forms of life? Does not the beaver build dams, the squirrel store nuts, and the spider spin webs?

To us, it would seem there is simply no comparison. In none of these cases, or any others involving the "lower animals," is anything like human forethought involved. The difference is so extreme that man must clearly be placed in a class by himself.

So, I repeat, it seems to us. Yet is it possible that this difference, so clear to us, is the result of mere prejudice and self-love? Would a creature from another world studying life on the earth really see the difference between a spider building a web and a man casting a net into the sea as anything more than one of degree?

It would be interesting (and soothing to our pride) if we could find some distinction between ourselves and all other creatures that is physical, concrete, and plainly visible even to the most alien of extraterrestrial creatures. We have begun by trying to make the distinction rest upon the manner of our utilization of energy as compared with other creatures. We can continue by exploring that point further.

One way of doing this is to start at the beginning, when man in the course of his evolution was *not* greatly removed from other animals, and try to find, as his culture developed, that point where the distinction was unmistakably set up, and see what that distinction was.

In doing so we shall be forced to explore further the significance of "energy" and "work." This, in turn, will lead us to a better understanding of the mechanism of life, which so far we have defined only as "something that makes an effort."

Let us ask, then, not for what purpose man expends energy (since this is hard to explain to an alien) but in what manner he does so.

To begin with, man, in expending energy to perform work, can do so only at a certain fixed maximum rate. If he were to attempt to lift twenty rocks each weighing fifty pounds and remove them to another spot a hundred feet away, he could do so. It would help him to move slowly and to rest between rocks so that the body could renew its energy supply at a rate equal to that at which it was expended. Still, by spacing his efforts judiciously and working slowly enough, a man could complete the task and, indeed, move any reasonable number of such rocks any reasonable distance.

If, however, the twenty fifty-pound rocks were united into one half-ton boulder, the man would not be able to lift it. He could not accumulate enough energy and expend it quickly enough to lift the half ton of rock at a stroke. The rate at which energy is expended is called "power" by sci-

entists. We might say then that, although a man has enough energy (given enough time) to lift a half-ton boulder, he does not possess sufficient power for the purpose.

In this respect man is like other animals, and enjoys no distinction. He possesses more power than a smaller animal such as a cat or a mouse, and less power than a larger animal such as a horse or an elephant. Any animal, however, no matter how large, can exert only a certain fixed power.

To be sure, creatures have evolved in such a way as to make use more efficiently of the energy at their disposal. For instance, to attempt to bite into the tough muscle of one's prey by means of the convergence of two flat gums would be useless. Equip those gums, however, with hard teeth ending in sharp points and the energy of the closing jaws (although no greater than before) is now concentrated in those hard points. The total effort is the same, but the effort per unit area of contact is much greater. At the point of contact there is therefore penetration. A tiger with fangs and claws becomes a fearful predator and a large horse, with more total energy at his disposal than at the tiger's, is an easy victim.

A living creature can be designed in many ways to make increasingly effective use of the energy at its disposal, so that rapid swimming, running, and even flying become possible. In every case, however, there remains a limit to the rate at which energy can be expended.

Where an organism is not already designed to make most efficient use of energy in some particular way, it is possible to make use of external objects ("tools") for the purpose. This is not a common habit among species other than man, but it is not unheard of. Birds may use rocks to break snail shells and even insects may use pebbles to block tunnels. Man and the various extinct manlike species so far surpassed all other creatures in the use of tools, however, that all the achievements of the lower animals in this respect can be ignored and man can be defined as a "tool-using animal."

A crucial step came when early man no longer made use merely of suitable objects he happened to find (as lower animals do) but set about consciously manufacturing stones with sharp edges and points.

In the sense that a tool may be a device whereby energy

applied at one point is transmitted, in more useful form, to another point or in another direction, that tool is a "machine." We usually think of machines as large, complicated structures, but for many thousands of years the machines used by mankind were simple indeed.

One of the simplest machines is a lever. Imagine one end of a beam inserted under a half-ton boulder and pivoted over a small rock nearby (see Fig. 1). A long extension of the beam stretches out beyond the pivot. If you lean downward at the end of that extension, the effort you exert is converted into a push upward at the other end of the beam (which acts in the same way as a seesaw—a kind of lever, after all—in which the movement of one child downward lifts the other child upward).

The change in direction of the push is enough to make the lever a machine, but there is more to it than that. The short end of the lever moves upward a lesser distance than the long end moves downward and the force¹ is correspondingly greater at the short end. At both ends the force times

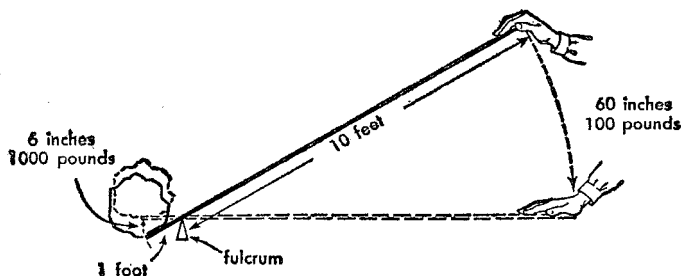


Figure 1. *The Lever*

the distance moved is the same. (This is the "principle of the lever" which was discovered first by Archimedes, a Greek mathematician who lived in the third century B.C., although, of course, mankind had been using the lever effectively, without knowing the principle, for many thousands of years previously.)

¹ The quantity of effort involved in a push or a pull may be used as a rough definition of the word "force." Force has a precise scientific meaning which we shall leave for later, accepting intuition for the time being.

If the long extension of the lever is ten times the length of the short extension, then as you push down sixty inches the rock would move up six inches, the difference in distance moved being in proportion to the difference in extension of the two parts of the lever. On the other hand, the difference in force exerted would be in reverse. You would have to exert only one tenth the force at the long end that was required to lift the rock at the short end. You would push down sixty inches with a force of a hundred pounds, in other words, to move the thousand-pound boulder upward six inches. This is predicted by Archimedes' principle, since 100 times 60 is equal to 1000 times 6.

The total energy expended at either end of the lever is the same, but by moving a hundred pounds over sixty inches you are expending energy at a rate within human capacity, whereas by directly moving a thousand pounds over six inches you would be requiring more of your body than it could deliver. The lever, therefore, makes it possible for you to lift singlehanded a boulder that, without the lever, could not be budged. (Archimedes expressed this in a famous sentence that goes: "Give me a place to stand on and I could move the world.")

Man, in the early dawn of his prehistory, devised other simple machines that served to reduce the necessary rate of expenditure of energy to a level within the capacity of the human body. Wheels were devised, as were pulleys, wedges, inclined planes, and so on.

In no case, however, does such a tool or machine increase the total energy at man's disposal. Although such devices, and nothing more, made it possible for the ancient Egyptians to build the Pyramids and the ancient Chinese to build the Great Wall, this was only at the price of condemning tens of thousands of individuals to backbreaking labor for decades.

Our alien observer, watching men swarm over the quarries and deserts of Egypt and the slow, painful stone-by-stone construction of the Great Pyramid, might not see much of a distinction between this and the construction of a termite hill. In fact the termite hill is larger compared to the insect than is the Great Pyramid compared to a man, and the alien might be more impressed by the termite. To be sure, the man manufactures his simple tools while the termite has his built into his body, but the alien might con-

sider this sign of greater ingenuity on the part of the man to be interesting but not crucial.

Man, despite all the tools he developed through tens of thousands of years of prehistory, remained as limited by the energy supply of his body as was the termite, and we have not yet arrived at a basic distinction between the two.

To perform more work in less time than can be performed by man and his hand tools requires a rate of energy expenditure greater than that which can be supplied by the human body. One way of attaining this involves the taming of animals.

The origin of animal domestication is lost in prehistory. It probably began as an accidental association, in which particular types of animals first skulked about campfires as scavengers, then were tamed for amusement and companionship, out of the affection that still binds small boys to turtles, beetles, hamsters, and other unlikely love objects.

And at last it must have occurred to men to breed these animals in captivity and to use the surplus numbers brought about through natural increase for various useful purposes: the meat for food, the pelts for clothing, the teeth and bones for ornaments and tools and so on.

In the same fashion, perhaps over the same general millennia, plants were, so to speak, also domesticated, and the art of agriculture had its beginning. The various groups of men, who began as hunters and food collectors, depending for their livelihood on the more or less random finding of food, now began to cultivate and breed their food, to become herdsmen and farmers. This meant that more human beings could be supported in a given area and, following the development of agriculture in particular, there came about man's first "population explosion."

In addition to all else, domesticated animals could be used as fellow workers. Those animals that were larger than men and could expend energy at a greater rate were particularly useful, especially since they could be fed on food coarser than that on which human beings could subsist and since, with proper treatment (e.g., castration), they could be made more docile and tractable than human slaves.

The most common draft animals of prehistory and of early civilized times were donkeys and oxen, and these represented an increase in available energy supply of two to seven times

over that available through use of an equal number of men. Horses can do better still, multiplying the available energy supply tenfold, but it was not until 1000 A.D. that the horse collar was invented. It was only after that that horses could be harnessed to pull loads without being choked by the effort.

However, if horses could not be used as draft animals, they could be used as mounts. Since the horse's body is more adapted for speed than our own is, a horse, even when carrying a man, can move more quickly than a man on foot can possibly do. In fact it was not until a little over a century ago that any form of land travel speedier than the man on horseback was developed.

Although the use of animals raises the rate at which energy can be expended, it is not really a very sharp rise, even allowing for the use of large animals such as camels and elephants. It is counterbalanced, partly, by the fact that the animals must be fed and cared for and that energy is expended in this fashion which might otherwise be used for man's more immediate needs.

Our alien observer, moreover, might see in our use of animals still another mere extension and elaboration of some of the activities of the lower animals.

To take an example, sea anemones are sometimes fixed by a crab to his shell. This is to the benefit of the sea anemone which, ordinarily quiescent, now finds itself transported from place to place, thus increasing its chance of finding food. This is also to the benefit of the crab since the stinging tentacles of the sea anemone keep creatures at a distance which might otherwise prey upon the crab. Moreover, the crab dines upon scraps of food that escape the sea anemone during the latter's meal.

Will our alien see a difference between this and the case of a man who rides a horse and feeds and protects the animal in return?

I think we do not, even yet, have a clear distinction between man and non-man.

But there is another way of securing a more rapid expenditure of energy than is possible for the human body unaided, and that is to turn to the inanimate world.

It may not be completely obvious that inanimate objects contain energy, but the matter isn't completely obscure

either. The wind and running water can move objects in a fashion that would certainly represent work if a human being were to duplicate the action. In fact, under stormy conditions, wind and water can, in short order, outdo the ravages of an army of men.

In the course of time man learned to make use of such inanimate sources of energy. Running water carries ships with their heavy loads downstream, while the energy of wind pushing against sail will carry those same ships over the sea in the absence of current. The energy of wind and water can be used to turn mill wheels, which will in turn grind grain to flour.

(Undoubtedly it puzzled early man to find such quantities of energy in inanimate nature when a superficial view, confusing energy with the ability to make an effort, made it seem a characteristic property of life, of the lizard rather than the rock. This dilemma was resolved by giving life to apparently inanimate nature and imagining the existence of demons or gods who made themselves manifest in the storm blast. The wind was their breath and the waters raged at their command.)

However, long before wind and water were even feebly tamed by man, another source of external energy came under his command. This is commonly referred to as the "discovery of fire" and was achieved one hundred thousand years ago or more, not by men of the modern type (who did not exist then) but by manlike beings of species now extinct.

Fire itself, of course, was never really "discovered," since it has always existed in the consciousness of sentient beings if only because lightning must every once in a while start a forest fire. Ordinarily such a forest fire means panic, flight, and possibly death to any living things in its path, but perhaps individual pre-men may have, on occasion, been attracted to the quietly burning remnants of such a fire and found amusement in playing with it, in feeding it more sticks, and in watching the dancing flames. (The fascination, sometimes deadly, which fire holds for children today is well known to all parents.)

If this happened at night the pre-man might appreciate the welcome light and warmth. He might even find that other animals kept their distance (being, unlike man, too sensible to approach fire) so that it meant protection too.

Eventually a man might take the risk of "domesticating"

fire by keeping such a flame on his campsite or in his cave in order to obtain all these benefits. After a time he would find that heating food over the fire (especially meat) made that food easier to chew and gave it a more pleasant taste.

The mere use of fire in such a way was not the key point of the discovery, however. No matter how well a fire is tended, it might inadvertently go out. Then what? There is, of course, the insurance of keeping a number of fires going so that one can be relit by means of a burning ember from another (or, for that matter, from the burning ember of a fire maintained at some other campsite or by some other tribe). This, however, evades the problem rather than solves it.

Clearly the key discovery would have been a method of deliberately kindling a fire where no fire existed before. If even a spark of fire could be directed into a heap of finely divided inflammable material, that would do. Such a spark might be struck if two rocks of the proper sort were sharply brought together; or else the friction of a twirling stick or the sudden compression of a volume of air could provide the necessary heat.

Whatever the method used, the gaining of the ability to start a fire at will marks the real "discovery of fire."

Now fire is a concentrated source of energy. With its use, the amount of energy at the disposal of a single human being is so much greater than is to be found in his own body that it can be considered virtually limitless. It is for this reason that the "discovery of fire" is beyond doubt the greatest single human achievement. It alone freed him completely from bondage to the limited energy supply of his own body, eked out by that of the animals he domesticated.

Here at last our unwearied alien observer can finally find his clear distinction. Here is an unmistakable physical achievement in the manner of utilizing energy, which marks out man (and pre-man) from all other animals. No other species, however intelligent, makes even the most fumbling attempt to use fire, whereas no tribe of man is known today, however primitive, that does not use fire.

Chapter 2

THE CONSEQUENCES OF FIRE

The existence of fire is far more than a means for handily distinguishing man from other creatures. The vast amount of energy placed at the disposal of man, through fire, could be, and was, used to revolutionize the nature of his existence.

Nor was man unmindful of his debt to fire. The Greeks had their myth of the Titan, Prometheus, saving the human race from misery by bringing down the celestial gift of fire from the sun. The sun, as the prototype of fire, was worshiped by many, notably by Ikhnaton, the Egyptian Pharaoh of the fourteenth century B.C. whose attempt to introduce a purified sun-worship did not survive him. In its earthly form, fire was worshiped by the Zoroastrians of ancient Persia, a religion surviving in modern times among the Parsees of India.

Among the Norsemen, fire seemed to be recognized in its malevolent aspect, for Loki, the Norse near-equivalent of Satan, was the god of fire. Among the Greeks and Romans, however, Hephaestus (Vulcan) and Hestia (Vesta) were connected with the beneficent fire of the forge and hearth, respectively. Sacred flames were tended and kept perpetually burning by the Vestal Virgins (perhaps a dim memory of the early time when a fire extinguished was a minor catastrophe).

Yet despite all this adoration, did many men, I wonder, suspect that in a proper understanding of the phenomenon of fire lay hidden an understanding of the very fundamentals of the universe and, in particular, an understanding of the phenomenon of life itself?

Although I shall now begin the story of fire by describing its effect upon the external affairs of men, I intend to arrive,

by this route, at a reconsideration of life in terms, so to speak, of fire.

The mere fact that fire was a source of light and heat independent of the sun meant that man could roam beyond the tropics that imprison his nearest living relatives, the great apes, and into the damp, cold regions with seasons of snow and long freezing nights. Most particularly, fire enabled him to withstand and survive the advancing glaciers of the ice ages.

In addition, the heat of the fire brought about changes in food that were the equivalent of partial digestion, and this made ordinarily inedible food palatable and nourishing. Man's food supply was thus multiplied greatly. The use of fire, by warding off the large predators, added further to man's security.

It was fire and fire alone that enabled man to become a creature native to all the world and put mastery into his hand. Nor has the importance of fire diminished with time; rather the reverse. Even in the course of the last few centuries—a hundred thousand years or more after the initial discovery—new fuels and new methods of fire ignition could still introduce new revolutions in man's way of life. Wood was undoubtedly the first fuel used in building and maintaining a fire. Coal took primacy of place in the 17th century, however, and petroleum in the 20th.

For at least 95 per cent of the time during which man and fire have been associated, the gifts of fire remained rather intangible, consisting chiefly of light and warmth. It was just before the beginning of historic times that the use of fire presented mankind with something completely new in the way of tangibles. He was introduced to the metals.

It was not just the knowledge that metal existed that was of importance, any more than was the mere knowledge that fire existed. Probably throughout man's stay on earth individuals had come across small nuggets of glistening material that did not behave as stone did. Stone, if beaten with another stone, shattered and powdered. This glistening material, on the other hand, was deformed, but remained coherent. It changed shape and could be beaten thin and into intricate forms. It was almost inevitable that such material would come to be valued for ornaments. Gold and

copper ornaments have indeed been found in Egyptian tombs dating far back to prehistoric times.

The crucial breakthrough took place with the discovery that metal could be obtained from material that did not in itself seem metallic. It was the discovery of a method of forming metal where no metal was previously evident that was important, just as many millennia earlier it had been the discovery of a method of producing fire where no fire had previously existed that was of importance.

Undoubtedly the discovery of a way of producing metal came about at first by accident. Copper attaches itself to other substances such as oxygen and sulfur rather loosely and, in the process, forms certain bluish rocks. If such rocks are heated strongly with carbon, the carbon removes the oxygen and sulfur from the copper and combines with them. The combination of carbon with oxygen and sulfur forms gases that vanish in the smoke of the fire. Left behind is the copper in its metallic state.

Carbon alone, mind you, was insufficient for the purpose of converting rock to metal. It had to be carbon plus the intense heat which, in the case of primitive man, could only be supplied by fire. Copper, and metals generally, are thus direct gifts of fire. In time, as the art of metallurgy progressed, first bronze (a copper-tin alloy), then iron, became of prime use in weapons and armor.

Imagine then a campfire built on copper-containing rocks ("copper ore") purely by chance. The carbon of the charring wood would do its work and once the fire was out some sharp-eyed shepherd might then notice copper nuggets among the ashes. What happened by accident (perhaps a number of times) could eventually be repeated on purpose and it might be borne in on men that fire could be used to, in effect, melt metal out of rock. By 3500 B.C. the deliberate smelting of copper ore was a well-established industry in the Near East.

Metallic ores are not exactly common, but they are certainly more common than are deposits of native metal and so, after 3500 B.C., metal for the first time came into reasonable supply. Copper could be used not only for jewelry in quantities that brought the ornaments within the reach of the common people, but also for utensils. The energy of fire made possible other advances, minor perhaps in comparison

with the importance of the discovery of metallurgy, but not so minor that we would willingly abandon them. The baking of clay produced pottery, porcelain, and bricks for a thousand purposes, while the strong heating of sand with soda and limestone produced glass for a thousand other purposes.

It is customary to think of the Middle Ages as a period of decline after the great days of the Graeco-Roman civilization, yet it was medieval Europe that produced astonishing advances through the development of sources of energy other than that of ordinary fire.

There was the introduction, for instance, of new varieties of combustion, and, as is often the case, it was the exigencies of warfare that brought this about.

Thus, the Greeks of the Byzantine Empire (the medieval remnant of the old Roman Empire), in the seventh century A.D., developed a combustible mixture, the secret of which has been lost. It burned on water and could not be put out by water. Such "Greek fire" was deadly against wooden ships, and the Moslem Arabs, besieging Constantinople, were repelled twice, once in the seventh century and again in the eighth, chiefly through the use of Greek fire against the Moslem fleet. Constantinople was thus given seven additional centuries of life as a Christian city before it finally fell to the Moslems (Turks, this time) in 1453.

Greek fire, like ordinary fuel, depended on air to burn (or, more precisely, on the oxygen in the air).

The true breakthrough came with the discovery of gunpowder when combustion was made independent of air. Somehow (again, no doubt, by accident to begin with) it was found that if combustibles such as charcoal and sulfur were mixed with saltpeter (potassium nitrate) and heated, a rapid and forceful burning would take place, even if air were excluded. This is so (as people in the Middle Ages did not know, but as we do now) because the potassium nitrate contains oxygen in loose combination, so that, in a sense, it is oxygen being mixed with the combustibles.

Primitive cannon were supposed to have been first used at the Battle of Crécy in 1346. Artillery became steadily more important in the centuries that followed; hand weapons using gunpowder were slowly developed; and all this changed the very structure of European society.

Another medieval advance of great consequence was the compass, which involved the use of a type of energy known as "magnetism" (from a Greek city named Magnesia, near which the first natural magnets were found by the Greeks).

It was the use of the compass, which gave navigators knowledge of directions when far from land, even when clouds hid the ordinarily guiding sun and stars, that encouraged the exploration of the ocean. It seems quite likely that without the compass the great voyages of Columbus and those who followed him would not have been made.

It might seem that the consequences of gunpowder and the compass are such that medieval times could offer no other discovery as important. And yet there is a third that is greater than either and one that, unlike the other two, does not directly involve any form of physical energy.

About 1450, just as the Middle Ages were coming to a close and modern times (often arbitrarily dated from Columbus' discovery of America) were about to begin, a German named Johann Gutenberg developed the technique of using movable type with which to print books.

No discovery prior to modern times caught on so quickly. Within a generation the flood of the printed word had begun. Learning became the province (at least potentially) of all. Scholars could communicate, not by the slow circulation of painfully handwritten treatises, but by the vast outpouring of printed volumes.

It was not merely the views of Copernicus that revolutionized astronomy in the next century, or the views of Vesalius that revolutionized anatomy. It was the fact that those views could be crystallized into books (published in the same year of 1543) in such quantity that they could penetrate every cranny of the European world of scholarship, spreading too quickly to be suppressed by ancient authority.

Printing led to an intellectual ferment, unleashed a kind of mental energy, if you will, that was a powerful driving force in the development of modern science. There was a new view of the universe, first fully demonstrated by the Italian scientist Galileo Galilei, which addressed itself to the accurate measurement of physical phenomena: the conversion of observation into quantitative and mathematically expressed relationships.

This led, inevitably, to the quantitative study of the phenomena of energy and work, so that the fire that had first made man something more than an animal a hundred thousand years before, came at last to be studied as well as used. Through such studies came new changes in man's way of life that dwarfed even those brought on by gunpowder and compass.

To begin with, the new age of science placed fire in a completely new position as help to man. For all the uses to which fire had been put in prehistoric times and in the ancient and medieval eras of historic times, its energy virtually never substituted, directly, for the energy of the human muscle.

Fire gave man altogether new things—metals, brick, glass—but without fire mankind would simply have had to do without, for there was no way in which the energy of muscle could have substituted for fire in this respect. And, despite these gifts of fire, man had to continue with all the old labor that was involved in living; he had to continue pulling, lifting, pushing, and chopping, with, for the most part, only the muscles of his domesticated animals to help his own.

In fact, in a number of ways, the occasional use of inanimate energy increased, rather than decreased, the necessity for muscular effort. Water, for example, flows from a higher level to a lower, and the greater the difference in level, the more rapid the flow and the more useful the energy content. The higher the rate of flow, the more efficiently a water wheel could be turned and a mill wheel powered. In England, however, many streams were relatively gently sloped and slow-moving. To increase the available energy, water had to be raised to a reservoir and allowed to fall from that, but raising the water meant the input of muscular work.

The particular problem that most concerned the English miners was that of raising water in order to get rid of it in the deep coal mines. The ingenuity of man had to turn to the problem of pumping.

Here, a vicious cycle was involved. To lift water requires energy and the energy could most conveniently be obtained from water already lifted. To break the cycle one had to depend on outside sources of power not involving falling water—either the erratic force of the wind driving the vanes

of a windmill, or the sharply limited energy of living muscle, either of man or of animals.

However, the very problem of the hand pump and the paradoxes to which it seemed to give rise suggested a new way of breaking the vicious cycle—one that was to free man of physical drudgery at last.

As mines grew deeper and pumps had to pump water higher, it became clear that water could not be pumped higher than thirty-three feet, however wildly the handle was yanked up and down.

Galileo tackled the problem of why suction pumps should reach such a limit but, for once, failed to see the solution. His pupil, Evangelista Torricelli, tackled the problem in his turn. Shortly after Galileo's death Torricelli speculated that water was sucked upward because of the weight of the air pressing down on the water reservoir below.

Ordinarily, Torricelli reasoned, air would press down equally on the water within the pump cylinder and upon the reservoir outside. When the piston was raised, however, and a vacuum created, the pressure was unbalanced. There would be full pressure on the reservoir and much less than full pressure on the water within the pump. The water within the pump would be pushed upward.

At a height of thirty-three feet, the column of water that had been built up by pumping action would weigh as much as the column of air pressing down on the reservoir. The two columns would balance and the water would be lifted no higher. A liquid lighter than water could be pumped to greater heights, if this were so, while one heavier than water could not attain even the height of thirty-three feet.

To test this theory conveniently, Torricelli made use of the densest liquid he knew, and therefore the one that could be lifted to the least height. This was mercury. In 1643, Torricelli poured the liquid metal into a three-foot tube of glass, closed at the bottom, and carefully upended it into a large dish of mercury. The full weight of the air pressing downward on the mercury reservoir was not sufficient to balance a three-foot column of mercury. The mercury poured out of the tube, therefore, until its height above the reservoir was thirty inches. Then it poured out no more.

The column of mercury remained balanced at that level, giving visible proof of the force represented by air pressure.

Above it was the first good vacuum (still called a "Torricellian vacuum") ever produced by man (see Fig. 2). Torricelli had invented the barometer, and it is still used today in essentially the same form, measuring the slight differences in air pressure from day to day and hour to hour in order that changing weather might be forecast.

Torricelli had thus demonstrated that air had weight and could therefore exert pressure. In fact, air pressure at sea level is some 14.7 pounds per square inch, so that the total pressure of the atmosphere against our bodies comes to about 22 tons. This is not evident to us because air exerts its pressure equally in all directions, from within our tissues outward as well as from outside inward, so that matters are in balance and the pressure goes unnoticed.

If anyone doubted the power of a man-made pressure difference, he need only have been present at demonstrations conducted by the German physicist Otto von Guericke (who

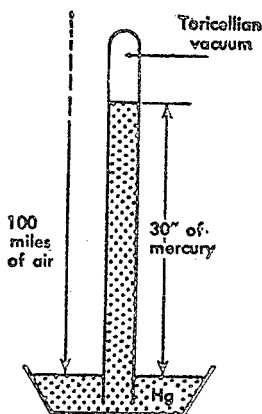


Figure 2. *The Barometer*

was also mayor of the city of Magdeburg). The work of Torricelli interested Von Guericke to the point where he devised a suction pump that would lift air rather than water. Attached to a closed container and powered by a hand crank, moving pistons trapped and removed bit of air after

bit of air until a fairly good vacuum was produced within the container.

Von Guericke put his air pump to use in the following fashion. He fitted two metal hemispheres together by a mere meeting of a horizontal flange at the edges, with no bolting or fastening whatsoever. One of the hemispheres was fitted with a one-way valve and by attaching his air pump to that valve he was able to create a vacuum within these so-called "Magdeburg hemispheres." Now it was air pressure holding them together, fourteen pounds or so pressing on each square inch. To demonstrate the force of that pressure, Von Guericke had a team of eight horses attached to each hemisphere and driven in opposite directions. The hemispheres held. When air was allowed to enter again, the hemispheres fell apart of their own weight. (If you affix a rubber suction cup to a smooth, moist surface and try to pull it off by main force, you will be repeating this experiment in a small way.)

In another demonstration, Von Guericke pumped a vacuum on one side of a well-fitted piston and in that way slowly pulled it down a cylinder against the attempts of fifty men, pulling on a rope, to keep the piston in place.

All this dramatized the power inherent in pressure differences, and yet the power was only a scientific curiosity. It could not be put to practical use.

What was really needed was to devise a method to produce the necessary pressure differences by some method better than muscle power. Muscles power the suction pump and Von Guericke's air pump (and for that matter, they power the lungs as well). To be sure, Von Guericke showed that one man on a hand-operated air pump could outdo fifty men on a rope, but the one man developed his victory over a long period of pumping. Industrial operations could not very well advance if power could be concentrated only in such slow motion.

The answer to this problem turned out to involve steam, the hot vapor into which water turned when it boiled, and that deserves a slight digression.

In 1690 it occurred to a French physicist named Denis Papin that, if water were boiled and the steam allowed to fill a container, that steam would displace most or all the

air in the container and would be, to begin with, at the same pressure as the air it displaced.

But note the difference. Air could be removed from the container only by effortful pumping. The steam that had replaced it could be removed by the simple expedient of cooling the container, for the large volume of steam would then be condensed to a few drops of water, leaving a virtual vacuum in the container. The whole process would involve, as the source of energy, nothing more than some burning wood that would serve to boil the water and form the steam in the first place. No muscular effort, no tedious pumping would be involved. In fact, where pumping would create the vacuum only after an extended period, a jet of cold water on the hot chamber would condense the steam and produce the vacuum in a matter of seconds. Gone would be the slow motion of the air pump.

Papin actually devised a machine so constructed that the vacuum produced forced a piston downward through the action of air pressure, this in turn lifting water upward. The device could not perform as a practical pump, but it did show that, in principle, water could be lifted by means of a steam-produced vacuum.

A few years later, in 1698, an English military engineer, Thomas Savery, invented the first practical device of this sort. He raised water as high as possible by means of a steam-produced vacuum, then put steam to use in another way. By building up steam pressure in a closed container he could inject a forcible jet of steam into the column of raised water and blow it up and out. The trouble with the Savery device was that the use of high steam pressures in containers built with the techniques of the late seventeenth century was dangerous. The possibility of explosion and the scalding to death of workers using such an engine was ever present. The Savery engine saw some use but was definitely not worth the risk.

In the next decade or so, however, an English blacksmith, Thomas Newcomen (who may not even have known of Savery's engine), devised a pump that made use of steam at atmospheric pressure only. A chamber was filled with steam and then cooled, filled with steam again and then cooled, over and over. In a way it was Papin's one-step device multiplied into continuing repetition, so that it duplicated the action of a pump being manipulated by hand, except that

the energy was coming from burning wood or coal rather than muscle.

No high-pressure steam was involved, and the comparative lack of danger made Newcomen's engine popular at mines where seeping water had to be pumped out. In 1778 more than seventy Newcomen engines were working in Cornwall alone.

Newcomen's engine was, however, still slow, and terribly inefficient besides. It used vast quantities of fuel for the work it did, wasting more than 99 per cent of the energy of the fuel. In the 1760s, the English civil engineer John Smeaton improved the engine and doubled its efficiency, which still left it low indeed. Moreover, the Newcomen engine lacked adaptability. It was useful only for the one activity of pumping.

Something more was needed and on the scene stepped James Watt.

In 1765, James Watt, a Scottish mechanical engineer and instrument maker, was given a Newcomen engine, which had broken down, to repair. Even after he repaired it he was dissatisfied with the way in which it functioned. During the course of a thoughtful Sunday walk it seemed to him that he perceived the chief source of inefficiency. The chamber, having been heated to steam temperature by its steam content, was then cooled to produce the vacuum. So far, so good. But then, in the next cycle, when steam was reintroduced into the chamber, the steam was condensed by the now cold walls. It was only after a considerable quantity of steam had been driven into the chamber that its temperature was raised to the point where the steam could remain as steam. At every cycle, then, immense quantities of fuel were required just to undo the work of the cold water that had cooled the chamber and produced the vacuum in the previous cycle.

Watt introduced a second chamber (a "condenser") into which the steam could be led. The condenser could be kept cold constantly while the first chamber (the "cylinder") was kept hot constantly. In this way the two processes of heating and cooling were not forced to cancel each other. By 1769, Watt had a working steam engine with an efficiency that was greatly increased over any of the Newcomen variety. Furthermore, since there was no long pause at each cycle

to heat up the chamber, Watt's engine did its work much more quickly.

Watt introduced many other ingenious improvements, such as that of allowing steam to enter alternately on either side of a piston. Previously air pressure had driven the piston rapidly in only one direction as the vacuum was produced, leaving it for mounting steam pressure to move it slowly back in the other direction. With steam on either side alternately, air pressure drove it rapidly in each direction and efficiency moved up further. By 1790 the Watt engine had completely replaced the older Newcomen engine and by 1800 some five hundred Watt engines were working in England.

Today, Watt is considered the "inventor of the steam engine," which of course, in a literal sense, he is not. The reason he is so considered, however, is not merely because he improved the Newcomen engine but because he was the first to make such an engine more than a pump. In the 1780s he devised mechanical attachments that ingeniously converted the back-and-forth movement of a piston into the rotary movement of a wheel, and by one movement or the other, the steam engine could be made to power a variety of activities. Early in the game, iron manufacturers were using it to operate bellows to keep the air blast going in their furnaces, and to power hammers that crushed the ore. The now versatile steam engine had become the first of the modern "prime movers"—the first modern device, that is, to take energy as it occurred in nature and apply it to the driving of machinery. With that was introduced the Industrial Revolution.

The steam engine has been continually improved in the direction of increased efficiency in the nearly two centuries since Watt's work; however, what brought the Industrial Revolution into the home was not the steam engine directly, but electricity.

The energy inherent in electricity, like that in steam, was known even to the ancients but was a plaything and a curiosity at first. By 1800 methods had been found whereby metals and chemicals could be combined in such a way as to produce a steady current of electricity. This current was of utmost importance to scientists but it could only be sup-

plied in small quantities and was far too expensive to be used on a large scale.

In 1831 the English scientist Michael Faraday discovered a new way of producing an electric current. He found that a copper disk, rotated by hand between the poles of a magnet, produced electricity continuously. That was the first "electric generator." It was then only necessary to keep a generator turning by water or steam power to produce electricity cheaply and in any quantity without the intervention of muscle.

About the same time the American physicist Joseph Henry devised a method whereby an electric current could be made to turn a wheel, and thus invented the "electric motor." Eventually, after several decades of improvements in the techniques involved, it was possible to supply the home with cheap electricity (thanks to Faraday) that could be used to power appliances by means of electric motors (thanks to Henry).

Toward the end of the nineteenth century it was found that the electric current consisted of particles that could be driven through vacuum and with an intensity and in a direction delicately controlled by electric fields. These particles ("electrons") gave birth to the electronic gadgets of today, notably to radio and television.

Anyone, then, who would compare the state of man in the latter half of the twentieth century with that in 1750, and then compare his state in 1750 with that in the Stone Age, might well come to the conclusion that, despite all the changes that took place before 1750, those that took place after 1750 were the more startling and radical.

The conclusion that seems most reasonable is that, of all the technological advances in man's history, those most pregnant of consequences were, first, the discovery of fire and, second, the invention of the steam engine. The first made the energy of combustion available to man, the second bent it to use as prime mover.

The history of man, then, it would seem to me, is much more a history of the development of his uses of energy than it is the story of the vagaries of kings and conquerors.

Chapter 3

THE MEASUREMENT OF MOTION

The new scientific attitude set in motion by Galileo somewhat before 1600 was not primarily dedicated to the development of practical improvements in man's use of energy. That might, in fact, be considered an almost accidental by-product.

Rather, scientists bent the new attitude of quantitative measurement to the purpose of sating their curiosity concerning the workings of the universe and the principles on which those workings were ordered. (If this produced useful and beneficial innovations, so much the better, of course.)

In the process they began with the quantitative study of various forms of energy—forms that had hitherto been used by man without real understanding, just as I have been speaking about them hitherto without definition. And what they discovered (although they had no inkling of it at first) applied as strongly to life and to the human body—as we shall see—as to the inanimate world they were so carefully analyzing.

One obvious form of energy is that involving movement, because it is movement that is most directly converted into what can be considered work. The moving shovel that bites into the soil, the moving ax that wedges into the tree trunk, the moving vehicles that transport loads, the moving battering ram that breaks down a city wall, all show unmistakably the connection between movement and work. (The energy involved in movement is called "kinetic energy" from a Greek word meaning "to move.")

In the 1580s, Galileo experimented with moving bodies. Since no decent clocks existed in those days, he had to improvise methods for marking short intervals of time (including water dripping out of a pot with a hole in the bottom, and the beating of his own pulse). These did not suffice

to measure the rate of movement of a freely falling body, so he investigated instead the motion of balls rolling down an inclined plane. By making the plane less and less steep, he could cause the body (starting from rest) to gain velocity as slowly as he pleased. He thus produced the effect of a freely falling body in slow motion.

The results of his investigations were systematized just a century later, in 1683, by the English scientist Isaac Newton, who presented the world with what are now known as the "three laws of motion."

The first law was based on Galileo's observation that when a smooth ball was made to roll along a smooth level surface its velocity decreased very slowly. The smoother the ball and surface, the slower the decrease in velocity.

Newton's first law of motion expresses this as follows: "Every body persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it."

But though a ball rolling on a horizontal surface continues moving at an unchanging velocity (ideally), this is not true of a ball rolling down an inclined plane, or of one falling freely. Such bodies, as Galileo showed, underwent an "acceleration." That is, their velocity increased steadily as they rolled or fell downward. Galileo showed, by actual measurement, that this velocity increased in direct proportion to the time elapsed after the ball had started moving, and that the distance it traversed was proportional to the square of the time elapsed.

Now to change an unchanging condition requires an effort. This, at least, is the common experience of mankind. To set an object into motion requires an effort. Similarly, to stop a moving object requires an effort. Anyone who has tried to throw or catch a baseball knows that an effort is required in either case, an exertion of strength or of force ("force" coming from a Latin word meaning "strength").

Any change, then, from the changelessness of the first law of motion, any increase in velocity, any decrease in velocity, any change in the direction of movement, is an acceleration and this must be due to something that is defined as a force.

For any given body the size of the force is directly proportional to the acceleration involved. It is the common experience of mankind that it requires more effort to catch a

fast baseball than a slow one, and more effort to throw a "fast ball" than a "slow ball." On the other hand, it is also the common experience of mankind that it is harder to set a heavy object to moving (or to stop it once moving) than to do the same for a light body. It is less painful to the hands to stop a ping-pong ball than a baseball traveling at the same speed. The property of a body that describes the extent to which it will be accelerated by a given force is called its "mass." (The idea of mass was first made clear by Newton.) The greater the mass, the smaller the acceleration produced by a given force.

Common experience may give us a general idea of what a particular force will do, but it was by considering the quantitative measurements performed by Galileo that Newton could make the second law of motion the definite statement that goes: "A force exerted on a body will produce an acceleration on the body directly proportional to the magnitude of the force and inversely proportional to the mass of the body." Mathematically, this can be expressed as $a = f/m$ or $f = ma$, where f symbolizes "force," a "acceleration," and m "mass."

The moon, although it moves continuously, does not move in a straight line, but in a curved path about the earth, so it does not really stand as an example of the first law of motion. Its velocity is constantly accelerated in the sense that the direction of its motion is constantly changing, so some force must be continuously exerted upon it. Since the change in its motion is always in the direction of the earth, it seems reasonable to assume that the source of the force is here on earth.

Newton was able to show that the same force that caused an apple to fall to the ground with a certain acceleration sufficed to explain exactly the acceleration imposed on the moon in its travels through space. This helped show that objects in the skies do not follow natural laws that are different from those on the earth, but that the same basic laws of nature rule everywhere in the universe (at least to the best of our current knowledge).

Newton speculated that the attractive force that pulled the apple to the ground and held the moon in its curved path about the earth were specific examples of a general situation whereby every object in the universe exerted an

attraction upon every other object in the universe (the "law of universal gravitation").

Newton made use of the laws of motion (including the third law, which states: "For every action, or force, there is an equal and opposite reaction, or counterforce"—the law that is exemplified every time a rocket is sent out into space) to show that such a universal force of gravitation between any two objects must vary directly as the product of their masses and inversely as the square of the distance from center to center of those objects:

$$f = \frac{mm'}{d^2}$$

Newton's law of universal gravitation was found to explain the motions of the various heavenly bodies wonderfully well. It set up such a beautifully simple and workable scheme of the universe that it irrevocably established modern science as an important intellectual achievement and forever freed mankind from overadulation of the accomplishments of the ancient Greek thinkers.

(In the twentieth century a new and more sophisticated view of the universe has grown up that better explains some of the phenomena discovered in the realm of the immensely small and the immensely vast. The whole concept of force and of gravitation—indeed of motion, space, and time—has changed. Nevertheless, for most ordinary purposes the Newtonian view works well enough.)

Once the notion of the measurement of motion was firmly fixed in the scientific consciousness, it followed that work ought to be measurable too. Since the application of a force to a body moved it out of its "natural state" under the influence of whatever other forces were already acting upon it, it seemed simplest to measure the amount of work done by multiplying the force by the distance the body is moved against whatever resistance there is to the motion.

The measurement of such things as force and work involves units that are not familiar to the layman, but they begin with units of mass, distance, and time, all of which are familiar. We all know, for instance, that distances are mea-

sured in such units as inches, feet, and miles, while masses are measured in such units as ounces, pounds, and tons.

("Mass" and "weight" are often confused because the same units are used for both and because, on the earth's surface, a unit of mass is defined in such a way that it is equal in size to a unit of weight. However, weight is the effect of the earth's pull upon an object and this decreases with distance from the earth's center. Mass is the resistance of an object to acceleration by a given force and this has nothing to do with earth's gravity. Consequently on the moon, which has only one sixth the surface gravity of the earth, an object we think of as weighing six pounds will have a weight of only one pound, but will still have a mass of six pounds. Although the object will seem so light, it will be just as hard to impose an acceleration upon it on the moon as it would have been on the earth.)

Scientists make universal use of the metric system, in which the basic unit of length is the "meter" (equal to about 1.1 yards) and the basic unit of mass is the "gram" (equal to about 1/28 of an ounce). One hundredth of a meter is a "centimeter" (equal to about 2/5 of an inch) while 1000 grams is a "kilogram" (equal to about 2.2 pounds). The unit of time is, both in the metric system and in our common system, the "second."

Now "velocity" is the rate of displacement, in a particular direction, of an object in space, during a given lapse of time. A body can move two miles in one minute, in which case its velocity is two miles per minute. A physicist, using the metric system, would express a velocity as so many centimeters per second or as so many meters per second. For convenience, this is abbreviated as *cm/sec* or *m/sec*.

An acceleration is a change in velocity with time. Suppose an object is increasing its velocity uniformly with time. At a given moment it is moving 1 *cm/sec*, but a second later it is moving 2 *cm/sec*, a second after that 3 *cm/sec* and so on. Its velocity is increasing by 1 *cm/sec* each second and we can therefore say that its acceleration is 1 centimeter per second per second, or 1 *cm/sec/sec*.

It is possible to deal with these units as though they actually represented the mathematical fractions they resemble. For instance, *cm/sec/sec* can be handled as though it were $a/b/b$. If the latter quantity is considered to be $a/b \div b$, then it is $a/b \times 1/b$ or a/b^2 . Analogously, 1 *cm/sec/sec*

can be written 1 cm/sec^2 (and read "one centimeter per second squared"). If you are measuring your distance in meters, an acceleration would be measured as so many m/sec^2 .

Now according to Newton's second law of motion, force is equal to mass times acceleration (a relationship I expressed earlier in the chapter, you may remember, as $f = ma$). We must therefore introduce units of mass now.

It is conventional among scientists to use grams (gm) as the unit of mass where centimeters are being used as the unit of distance and kilograms (kg) as the unit of mass where meters are being used as the unit of distance.

If, then, the acceleration is being measured as so many cm/sec^2 , this is multiplied by the mass, in grams, of the body being accelerated, to give the force that must be applied. The force is therefore measured as gm-cm/sec^2 . If the acceleration is being measured as m/sec^2 , then the mass is measured in kilograms, and the unit of force is given in kg-m/sec^2 .

As a matter of shorthand, physicists have given the name "dyne" (from a Greek word for "force") to the combination of units gm-cm/sec^2 . One dyne can be defined, then, as that quantity of force which, imposed upon a mass of one gram, will impart to it an acceleration of one centimeter per second per second. Or, briefly, $1 \text{ dyne} = 1 \text{ gm-cm/sec}^2$.

As for the unit kg-m/sec^2 , that is given the name "newton." Since a kilogram is equal to 1000 grams and a meter is equal to 100 centimeters, a kilogram-meter is equal to 1000×100 , or 100,000 gram-centimeters. It follows, then, that 1 kg-m/sec^2 is equal to 100,000 gm-cm/sec^2 or that, in briefest form, 1 newton equals 100,000 dynes.

Next, consider that work represents a force exerted through a distance. If we consider the force as being measured in dynes, or gm-cm/sec^2 , and multiply that by distance in centimeters, we find that we are measuring work in units of $\text{gm-cm}^2/\text{sec}^2$. This unit is named the "erg" by physicists, from the Greek word for "work."

In other words, you do one erg of work when you exert a force of one dyne upon a body through a distance of one centimeter. Or, in symbols, $1 \text{ erg} = 1 \text{ dyne-centimeter} = 1 \text{ gm-cm}^2/\text{sec}^2$.

On the other hand, if force is being measured as newtons, or kg-m/sec^2 , that must be multiplied by distance in meters

to give a unit of work, which then turns out to be $kg\cdot m^2/sec^2$. This unit is named the "joule" (pronounced "jowl" in honor of a physicist I shall mention later).

One joule of work involves the exertion of a force of one newton through a distance of one meter. Or, in symbols, $1 \text{ joule} = 1 \text{ newton-meter} = 1 \text{ } kg\cdot m^2/sec^2$.

Since a newton is equal to 100,000 dynes and a meter is equal to 100 centimeters, a newton-meter is equal to 10,000,000 dyne-centimeters, or, which is the same thing, 1 joule is equal to 10,000,000 ergs.

With such phenomena as force and energy reduced to quantitative terms, the way was open for the discovery not only of the exact manner of change of certain phenomena in nature, but also of the manner in which other phenomena did not change.

It is apparent that a moving body can do work and therefore contains "kinetic energy" ("energy of motion"). From the laws of motion and from the definitions of force and work, it can be shown that the kinetic energy of a body can be represented by the following equation:

$$E_k = \frac{1}{2}mv^2$$

where E_k represents kinetic energy, m represents mass, and v represents velocity.

Now suppose a 10-kilogram (about 22 pounds) iron sphere is hurled upward at a velocity of 49 meters per second (160 feet per second). At the moment it is shot upward its kinetic energy is $(\frac{1}{2})(10)(49)(49)$ or just about 12,000 joules.

However, as the sphere travels upward, the force of gravity pulling down upon it gradually slows its velocity. After one second it is moving upward at a rate of 39.2 meters per second and after two seconds it is moving upward at a rate of only 29.4 meters per second. After five seconds (at which time it has reached a height of 122 meters) it will have been slowed to zero velocity (at least with respect to the surface of the earth). For an instant of time it will hover at that height in a state of rest.

But as the velocity of the object decreases, so does its kinetic energy. At the moment it becomes motionless, its kinetic energy is $(\frac{1}{2})(10)(0)(0)$ which comes out to zero, of course.

What happens to the kinetic energy that thus steadily diminishes and finally disappears as the object climbs? Is it gone forever? The answer is no, for the iron sphere, having reached the top of its climb, now begins to fall; slowly at first, then more and more rapidly under the pull of gravity. When it returns to the surface of the earth (and we are ignoring any effect of air resistance at this point) it has precisely the same speed as it had when it left the ground, except that it is moving downward rather than upward.

Once again it is moving at the rate of 49 meters per second and once again it has a kinetic energy of 12,000 joules. All the kinetic energy with which it had started had returned at the end. Where was it, then, at the top of the flight?

To answer that question, consider the fact that an object may possess energy merely by virtue of its position. A stone at the top of a cliff does not seem to possess particular energy while it lies there. It can lie there forever without doing any work. If that same stone were dropped off the cliff, it would gain velocity (hence kinetic energy) as it fell, and if it struck a man at the bottom of the cliff, it would kill him. Energy by virtue of position is called "potential energy" (which we can symbolize as E_p) because it does not seem to be there until you do something to give it a chance to appear, so that it is a potential property, rather than one actually in being.

The iron sphere (or stone, or any mass, in fact) loses kinetic energy as it rises but gains potential energy. At the very top of its elevation, all of its kinetic energy has been converted to potential energy. Then, as it falls, the potential energy begins to be converted to kinetic energy again and when it hits bottom its energy is all in the form of kinetic energy again. The potential energy at the top of its climb is just equal to the kinetic energy at the bottom. If, in fact, the sphere had been dropped from some intermediate height, it would develop a kinetic energy by the time it reached the surface of the earth that was just equal to the kinetic energy it had lost in reaching that height. At any point in its flight, then, its loss of kinetic energy is balanced exactly by gain in potential energy and vice versa.

The energy has not, after all, either disappeared or reappeared. It is there all the time, merely shifting from one form to another. The energy is "conserved." If we lump kinetic energy and potential energy together as "mechanical

energy" (E_m) then we can say that this expresses the law of "conservation of mechanical energy."

Here is one important glimpse of a kind of constancy in nature that could not have been appreciated until men actually started measuring phenomena accurately.

Of course, we mustn't think that, in actual fact, *all* the potential energy has been converted to kinetic energy when an object falls and reaches the earth's surface. It may fall no more and do no more work, but if we dig a pit a thousand feet deep, the fallen object, which had apparently come to the end of its fall, can suddenly fall again and do more work. Has additional potential energy appeared out of nowhere?

No, it is a matter of reference. An object at the top of a cliff has so much potential energy with respect to the surface of the earth, but it has much more potential energy than that with respect to the center of the earth; for if a hole deep enough existed, the object could fall four thousand additional miles after reaching the surface of the earth. An object even at the center of the earth still has potential energy with respect to the sun, and even at the center of the sun has potential energy with respect to the center of the galaxy. Nowhere in the universe is potential energy truly and absolutely zero, and in fact this lack of an "absolute frame of reference" is one of the central points of Einstein's theory of relativity.

This lack of an absolute reference need not matter to us, however. Physicists are interested in differences of potential energy rather than in absolute values. That is, an object on top of a mile-high cliff has a greater amount of potential energy than one at the bottom of the cliff and it is this difference in potential energy that is fixed as far as we, confined to our planet, are concerned. For mathematical convenience we can arbitrarily define the potential energy at the earth's surface as zero, and then the potential energy at any height becomes equal to the difference in potential energy at that height and the potential energy at the surface, since $x - 0 = x$.

Kinetic energy can be treated in the same way, as a difference rather than as an absolute value, and it turns out that mechanical energy is conserved regardless of the frame of reference.

Another interesting thing about potential energy is that

it depends only on position and not on the route by which the position is reached. Imagine three cannon balls of identical weight at the top of a cliff. One has got there by being fired out of a cannon in such a way that it just reached the top at the height of its climb. The second was treated by no means so violently but was brought to the top of the cliff by a long, winding path, sloping gently upward. The third was carried by an airplane twenty miles high in the air and lowered to the cliff by means of a long cable.

Now if those three cannon balls are dropped off the cliff, do you imagine that the one brought up gently would possess less energy than the other two? Or that the one which had temporarily gained immense potential energy while at a height of twenty miles would possess more energy?

No, all three would possess the same potential energy: that due to their position at the moment their energy is being considered. Their past histories have no influence.

In the same way, suppose one of the balls was simply dropped, while a second was allowed to slide down a long, frictionless incline sloping gently downward. The second would still reach the bottom with its full kinetic energy just as though it had been dropped. It would accelerate much more slowly on the inclined plane, but on the other hand it would accelerate over a much longer time. By the time the second ball reached the bottom of the incline it would have the same speed and the same energy as the ball that had dropped freely. And if the third ball had swooped down a frictionless roller coaster so that occasionally it interrupted its fall with a partial rise, it would still reach the bottom with the same kinetic energy.

Whenever a particular property of an object is conserved, as is the case with mechanical energy, then, when that object changes from state *A* to state *B*, the change in the conserved property depends only on the nature of states *A* and *B* and does not depend on the route taken in passing from state *A* to state *B*.

A physicist finds a satisfying neatness about a property that is conserved and is always on the lookout for examples.

Consider two objects that are inelastic; that is, objects that will not bounce away if they strike each other but will go "splat" and stick together. Soft wax or clay would be examples of inelastic substances.

Now imagine two wax balls of equal size approaching each other head on at identical velocities of 2 meters per second on some level, frictionless surface. They meet, squash against each other, and come to a dead halt. The two velocities have apparently canceled each other.

If the velocity of one ball is considered $+2$ m/sec, then the velocity of the other (moving in the opposite direction) would naturally be -2 m/sec. These two velocities add up to zero before collision, and, of course, add up to zero after collision.

If you were considering only this observation, you might begin to hope that velocity under such conditions would be a conserved property; that the algebraic sum of the velocities of component parts of a system might remain the same regardless of the manner in which the parts of the system change relationships among themselves.

But suppose the two balls of wax, just discussed, are not of equal size. Suppose the one traveling to the left is three times as heavy as the one traveling to the right but that both are still traveling at the speed of 2 m/sec. The total velocity is still zero, but when they collide and stick together they don't come to rest. The combined balls continue to move in the direction in which the larger ball had originally been moving, but at only 1 m/sec. The total velocity of the system is not conserved.

It was in 1671 that the English physicist John Wallis pointed out that what was conserved in these cases was not velocity alone, but mass times velocity (mv), this product being called "momentum."

In the case of the unequal balls, for instance, suppose that one ball weighed 2 grams and the other 6 grams. If they are each traveling at a velocity of 2 m/sec as they approach, then although their velocities are equal (except for the difference in direction of motion) their momenta are not. The smaller ball has a momentum of 2 times 2 or 4 gram-meters per second (gm-m/sec), while the larger ball has a momentum of 6 times 2 or 12 gm-m/sec. If the momentum of the larger ball is considered to be positive, the other, moving in the opposite direction, is negative and the total momentum of the system, before collision, is $+12 - 4$ or $+8$ gm-m/sec.

After collision, the combined balls have a mass of 8 grams and are moving at 1 m/sec in the direction that the larger

ball was moving originally, so the momentum is 8 times 1 and positive, or $+8$ gm-m/sec.

Experiment shows that momentum is always conserved under such conditions. Momentum is also conserved if two elastic balls (such as those made of glass or steel, which bounce rather than stick) collide and fly apart, provided we remember that, if velocity and momentum in one direction is taken as positive, then, in the other direction, it must be taken as negative. There are even recognized geometrical methods for dividing motions in any direction into positive and negative components, and momentum will be conserved even if balls collide other than head on, or if a number of balls are involved, as when the cue ball strikes the racked-up billiard balls in a game of pool.

Again, if a loaded rifle with a hair trigger is suspended from a cord and allowed to remain at rest, its momentum is zero. If the trigger is tripped and the rifle fires, a part of the system (the bullet) emerges from the muzzle with velocity and considerable momentum. The only way in which the system can retain its total momentum of zero is for the rest of the rifle to move backward, with a momentum equal and opposite to that of the bullet. Since the rifle is many times more massive than the bullet, it need kick back much more slowly to balance the momentum.

Angular momentum (mass times circular motion) is also conserved. This is most beautifully and gracefully demonstrated by a skater twirling as fast as she can. Each part of her body possesses a certain angular momentum, depending on the velocity with which that part is circling. Those parts of the body farthest from the central axis of the skater have the most angular momentum per unit mass because they are making the largest circles and are therefore moving most quickly. In the case of the skater, these parts are the extended arms.

If the skater moves her arms closer to her body, they lose angular momentum because they now make smaller circles per revolution and therefore move more slowly. But angular momentum must be conserved. What one part of the system loses, the rest must gain. Therefore, as the skater draws in her arms, her body as a whole speeds its revolutions and she whirls like a top.

Unfortunately, the conservation of mechanical energy and

of momentum only holds exactly in an ideal situation where air resistance and friction do not exist. Such a situation is never true on earth. In an ideal resistanceless, frictionless environment, a pendulum would never stop swinging, a bouncing ball would never stop bouncing, moving and colliding billiard balls would never stop moving and colliding, and a whirling skater would never stop whirling unless she made an effort.

None of this happens. When a ball is thrown into the air, not all its kinetic energy is turned to potential energy; some is lost in overcoming air resistance. Again, when it starts to drop, not all the potential energy that remains is converted to kinetic energy. Again some is lost as friction. And if it bounces and starts moving up again, some energy is lost in the process of the bounce because a ball is never perfectly elastic. No system, however cleverly designed, can overcome such losses in actual practice and so mechanical energy and momentum are never truly conserved.

Furthermore, friction makes it important after all to consider the route by which mechanical energy or momentum passes from state *A* to state *B*. A body dropped over a cliff and down through the air loses a relatively small amount of energy through air resistance. An identical body allowed to slide down an incline to the foot of the cliff (not an ideal frictionless one such as I imagined before but an actual one that does involve friction) will arrive at the bottom with considerably less kinetic energy than the body in free fall. The gentler the incline and the longer the distance of the roll, the greater the total friction and the smaller the final kinetic energy.

To bring about a truly conserving situation, frictional losses and other imperfections would have to be dealt with somehow, but for over a century and a half after Newton's time the matter was not handled adequately.

The missing factor involved heat, and scientists did not seriously study heat until after the steam engine made dramatic capital of the fact that heat was a form of energy that could be converted into work as truly and unmistakably as kinetic energy could. And it was only with the proper understanding of heat that physics could be made to deal with the actual world (and with life) and not with merely a fictitious world in the mind of the physicist.

Chapter 4

THE MEASUREMENT OF HEAT

Exactly what heat was, except that it was something to which the human body was sensitive, eluded the early scientists, but that did not prevent them from studying it as a phenomenon, and from achieving considerable success, too.

First, they learned to measure the intensity of heat; that is, the "temperature." Our sense perceptions give us rough ideas of the intensity of heat rather than of its total quantity. We can tell whether an object is hot or cold, but not how much heat it has. For instance, a bathtub full of cool water contains far more heat, all told, than does a drop of boiling-hot water. Yet the bathtub water feels cool to us while the drop feels hot. The intensity of heat is greater in the second case; more heat is contained in a given volume.

This is not only important to our senses but has much significance in the universe as a whole, for heat flows from the hotter object (higher temperature) to the colder one (lower temperature), and not necessarily from an object of greater total heat content to one of less. For instance, if the drop of boiling water is added to the cool bathtub water, heat flows from that drop (which cools off almost at once) to the bathtub water (which warms up very slightly).

If you're not certain you get the significance of this, consider the analogy to the flow of water. A mountain spring contains far less water than the ocean, but the spring is at a greater height; it has a higher potential energy than the ocean. Water does not flow from the water-rich ocean to the small water-poor spring. Instead it flows from the small mountain-high spring to the mighty sea-level ocean.

We even make a piece of folk knowledge out of the fact that "water always runs downhill," which is just the ordinary

way of saying that spontaneous motion is always in the direction of decreasing potential energy.

Temperature is to heat what potential energy is to mass. Heat flowing from high temperature to low is like water flowing downhill. (In fact, I wish there were some way of dropping the word "temperature" and using "heat potential" instead. There would be precedent because the driving force behind an electric current is called the "electric potential" and electricity always flows from a point of high potential to one of low.)

Now we can determine potential energy by measuring height above sea level. Is there any way we can measure temperature? Galileo was the first to see the correct principle involved. Almost all substances expand in volume as temperature rises and contract as temperature falls. In 1592, Galileo devised an "air thermometer" in which air changed volume with temperature and altered the level of water in a tube. Unfortunately, the water level was also affected by air pressure, so that results were erratic.

This was followed by sealed thermometers that were unaffected by air pressure. In such thermometers, liquid filled a bulb and partly filled a very thin, evacuated capillary tube. A very small change in volume brought about a comparatively large change in the capillary level of the liquid. The change in the level of the liquid could then be used to indicate the temperature.

In 1715 the German physicist Gabriel Daniel Fahrenheit first used mercury in a carefully graduated thermometer and designed the instrument approximately as we have it today. In the United States, Great Britain, Canada, South Africa, Australia, and New Zealand the "Fahrenheit scale," devised by Fahrenheit himself, is still used. On that scale, the temperature of melting ice is set at 32 degrees (generally abbreviated as 32°) and of boiling water at 212°. The use of the scale is indicated by the initial F. after the degree value.

In other nations and among scientists everywhere, use is made of a decimal scale devised in 1742 by the Swedish astronomer Anders Celsius, in which the temperature of melting ice is 0° and that of boiling water 100°. This is the "Centigrade" ("hundred steps") or "Celsius scale" and in either case this scale is indicated by the initial C. after the degree value.

Although the Fahrenheit scale is more familiar in everyday life here in America, I shall use the Centigrade scale in preference throughout this book, just as I shall make use of the metric system.

The use of the thermometer made it possible to tell something not only about temperature but about the amount of heat present in substances as well.

Suppose a liter of alcohol (a "liter" is a unit of volume in the metric system and is equal to about 1.05 quarts) is heated to 60° C. and that this is then quickly mixed with a liter of water at 20° C. Heat, naturally, will flow from the hot alcohol to the cold water. Our senses will tell us this since the mixture will have a temperature that we can easily judge to be intermediate between the hot alcohol and the cold water. That would seem logical, and without a thermometer there would seem no reason to investigate the matter further.

Once we have become accustomed to measuring temperature, however, we would expect more than merely "intermediate" as a description of the final state of the mixture. We would expect a half-and-half mixture of a 60° C. liquid and a 20° C. liquid to have a temperature of 40° C.—but in this case it doesn't. The final temperature of the alcohol-water mixture is only about 36° C.

Why is this? Apparently the amount of heat contained by alcohol at a given temperature is not as great as the amount of heat contained by water at that same temperature. To put it another way, the amount of heat that suffices to drive the temperature of alcohol up ten degrees is less than that required to drive the temperature of water up ten degrees. Conversely, the heat given off by alcohol dropping ten degrees in temperature is not as much as that given off by water dropping the same ten degrees in temperature. (It is as though water and alcohol represented a wide and a narrow cylinder respectively. It would take more liquid to fill a wide cylinder to a height of six inches than to fill a narrow cylinder to that same height. The two cylinders would be filled to the same height but the wide one would contain more liquid. In the same way, alcohol and water might be heated to the same temperature but the water would contain more heat.)

Now the temperature of the alcohol-water mixture is no

longer puzzling. When the alcohol cools off by 24 Centigrade degrees (from 60° C. to 36° C.) the heat liberated just suffices to warm an equal quantity of water by only 16 Centigrade degrees (from 20° C. to 36° C.) and 36° C. is the final temperature.

The Scottish chemist Joseph Black was the first to study this phenomenon. By 1760 he had made measurements of the amount of heat required to raise the temperature of a fixed mass of a number of substances through a fixed temperature range. He termed this quantity the "specific heat."

The specific heat of water is arbitrarily set at 1.00. It takes two thirds as much heat to drive the temperature of alcohol up ten degrees as it does water. The specific heat of alcohol is therefore roughly 0.65. The specific heat of another liquid called acetone is just about 0.50, and so on.

It turns out that the specific heat of water is quite high; indeed, among the highest known. For most metals, it is quite low. The specific heat of aluminum is 0.22; of copper, 0.093; of lead, 0.031. As a matter of common knowledge, we know how little time on the gas burner is required to turn an empty metal kettle dangerously hot, and how long a time it takes, in comparison, to warm even a moderate amount of water.

(In stating the specific heat of a substance one should, to be precise, state the temperature at which the value is determined, for the specific heat varies with temperature level. A hot substance generally has a higher specific heat than that same substance when cool.)

All this led naturally to a unit for measuring heat content (that is, the total quantity of heat contained by a given portion of matter; this is not to be confused with temperature, which is heat intensity).

The amount of heat required to raise the temperature of one gram of water from 14.5° C. to 15.5° C. was set equal to a "calorie," from a Latin word for "heat." (The final "ie," which looks a little unnatural, stems from the fact that the unit was introduced by French physicists who used the spelling conventions of the French language. The word is, on occasion, anglicized to "calory" by English-speaking people.)

A second and larger unit of heat content is that amount of heat required to heat one kilogram (a thousand grams) of water from a temperature of 14.5° C. to 15.5° C. Nat-

urally, this unit is a thousand times as large as the first and is, therefore, properly termed a "kilocalorie."

Confusion often results because the larger unit is sometimes carelessly called a calorie, with no distinction made except, possibly, to capitalize it thus: Calorie. A capital may serve to distinguish one calorie from another in reading, but certainly not in speaking. The worst offenders are the dietitians, who invariably speak of "calories" and really mean "kilocalories."

British and American engineers use a unit of heat similar to the calorie but based on the Fahrenheit scale of temperature. The amount of heat required to raise the temperature of one pound of water from 63° F. to 64° F. is called a "British thermal unit," commonly abbreviated B.T.U. A British thermal unit is equal to about 252 calories.

But to return to Joseph Black. He made another important advance in the study of heat. He observed that heat added to ice melted it but did not raise the temperature of the ice-water mixture until all the ice was melted. The amount of heat absorbed in the melting of a quantity of ice would, if applied to that same quantity of ice-cold water, raise its temperature nearly to the boiling point. In fact, it takes 80 calories to melt one gram of ice. Black referred to this as "latent heat," the word "latent" meaning "hidden" or "invisible," since the heat did not show up as a rise in temperature. Like the "potential" in "potential energy," the reference was to something that was there without seeming to be there. In modern physics, this form of latent heat is usually called "heat of fusion" ("fusion" being synonymous, in this case, with "melting").

Another form of latent heat is that which is involved in the boiling of a liquid. As heat pours into boiling liquid, there is no temperature rise until all the liquid is boiled. Here even more heat is involved, generally, than in the melting of a solid. It takes about 540 calories to vaporize a gram of water at 100° C. to a gram of water vapor that is still at 100° C. This is "heat of vaporization."

Black also found that when water was recongealed to ice (or when water vapor was recondensed to water) the latent heat that had been absorbed in the reverse process was given off again. Heat went into the substance when ice melted, and out again when water froze. It went into the substance when water boiled and out again when steam condensed. It was

like the shift from kinetic energy to potential energy and back again and pointed in the direction of heat as a conserved phenomenon or, at the very least, as being involved in some conserved phenomenon.

If change in temperature is measured by change in the volume of a liquid, then that volume change should be uniform. Otherwise, we would not be justified in marking off the degrees on a thermometric scale equally. The uniformity of volume change is by no means to be taken for granted. Water does not expand uniformly as temperature rises, for instance. In fact, between 0° and 4° C., water actually contracts with rising temperature. Mercury, however, was shown to change volume quite uniformly with rising temperature and that is one of the reasons it can be successfully used in thermometers.

Such investigations were, at first, confined to those liquids that were used in thermometers, naturally. Galileo's abortive attempt to use a gaseous thermometer seemed forgotten and for nearly two centuries thereafter very little work was done on the effect of temperature change on gas volume.

Thus, in 1660, the British physicist Robert Boyle first showed that the pressure and volume of a given quantity of gas were inversely proportional ("Boyle's law"). If pressure on a gas is doubled, the gas volume is halved; if pressure is tripled, volume is reduced to one third, and so on. This is one of the foundation stones of modern chemistry, yet Boyle never made the point that this relationship only held if temperature were maintained unchanged. Temperature and its effect on the volume of gases did not enter into his thinking, apparently.

About 1680 a French physicist, Edme Mariotte, rediscovered Boyle's law independently, and he did mention, incidentally, that temperature change also changed the volume of the gas. (Boyle's law is called "Mariotte's law" through much of Europe.)

Another step was taken in 1699 by the French physicist Guillaume Amontons, who noticed that different masses of air expanded by the same fraction of their volume with a fixed increase in temperature.

It was not until 1787, however, that the matter was reduced to exact measurement. The French physicist Jacques Alexandre César Charles finally showed that gases expanded and contracted uniformly with temperature change ("Charles'

law"—enunciated two centuries after Galileo had tacitly assumed the existence of the law by constructing his thermometer).

Charles found that a gas (and during the eighteenth century the existence of gases other than air had been established for the first time) expanded just $1/273$ of its volume at 0°C . for each Celsius degree rise in temperature. It correspondingly contracted by $1/273$ of its volume at 0°C . for each Celsius degree fall in temperature.

This means that if we begin with 273 liters of a gas at 0°C . and increase the temperature, the volume expands to 274 liters at 1°C ., 275 liters at 2°C ., 276 liters at 3°C ., and so on. In the reverse direction, the gas contracts to 272 liters at -1°C ., 271 liters at -2°C ., 270 liters at -3°C ., and so on.

It seems pretty clear that if Charles' law held exactly, then at -273°C . the volume of the gas would have contracted to zero and that at still lower temperatures the gas volume would have to have negative values, which is unthinkable.

Physicists were not too overwhelmed by this paradox because it depended entirely on the assumption that Charles' law continued to hold as temperature dropped to very low values and there were signs that it did not. In Charles' time and for nearly a century after, it was impossible to attain really low temperatures in order to test the consequences of this gaseous contraction. By 1900, however, it was quite clear that before -273°C . was reached all gases would have condensed to liquids, and Charles' law didn't hold for liquids. So much for the paradox.

Nevertheless, -273°C . would seem to be a sort of crucial temperature and some investigators began to think of it as an "absolute zero" of temperature, a kind of ultimate cold. It actually proved convenient to think of -273°C . (the best modern value is -273.16°C ., but we can ignore the fraction) as that temperature at which a system contains no heat at all. If we set that temperature equal to zero and work upward by the usual Celsius degrees, then the freezing point of water becomes 273° , the boiling point 373° , and so on. This is the "absolute scale" of temperature (see Figure 3).

In 1848 the British physicist William Thomson (who later was raised to the peerage and is therefore better known as Lord Kelvin) established this system on a firm theoretical

basis. For that reason, the absolute scale is often known as the "Kelvin scale" and the freezing point of water on that scale can be written either 273° A. (for "absolute") or 273° K. (for Kelvin). The latter is more common.

On the Fahrenheit scale, absolute zero is -460° F. (-459.69° F., to be more exact) and one could start at absolute zero and count upward by Fahrenheit degrees. The first to use such a scale was the Scottish engineer William John Macquorn Rankine, and this is called the "Rankine scale" in consequence. The freezing point of water on this scale would be 492° Rank. and the boiling point of water would be 672° Rank.

Lead melts	262°	327°	600°	620°	1080°
Water boils	80°	100°	373°	212°	672°
Body temp.	30°	37°	310°	99°	591°
Ice melts	0°	0°	273°	32°	492°
Oxygen boils	-146°	-183°	90°	-298°	162°
Absolute zero	-218°	-273°	0°	-460°	0°
	Réaumur	Celsius	Kelvin	Fahrenheit	Rankine

Figure 3. *Temperature Scales*

(The abbreviation is not "R." as might be expected because there is a Réaumur scale of temperatures, devised by the French physicist René Antoine Ferchault de Réaumur in the eighteenth century, and that is the one abbreviated "R." On the Réaumur scale, the freezing point of water is 0° R. and the boiling point 80° R. The Réaumur scale was never very popular and is only of historical interest now.)

As temperature and heat content became easily measurable phenomena, scientists welcomed them, so to speak, into

the fold and began to take a new and broader view of energy. With the coming of the nineteenth century, energy was no longer tied so entirely to motion. The concentration on mechanical energy had resulted in wonderful advances in the seventeenth and eighteenth centuries, but something more was now needed.

The steam engine had made it quite obvious that heat could do work just as surely as a speeding projectile could, and that heat was therefore a form of energy. Magnetism was clearly a form of energy, as was dimly recognized even in medieval times, and eighteenth-century experiments with electricity showed that to be a form of energy too.

The very word "energy" was introduced to science by the English physicist Thomas Young in 1807; and his most important investigations involved light, still another form of energy. In the next few decades the term broadened to its present extent, meaning anything that could produce work.

It began to occur to a number of investigators that the conservation of mechanical energy did not hold perfectly true simply because no account was taken of other forms of energy, particularly of energy in the form of heat. It had become quite clear, for instance, that the frictional losses that had to be subtracted from mechanical energy also resulted in the production of heat. Could the loss of mechanical energy be just balanced by the production of heat energy?

The outstanding experimentalist in this respect was an English physicist, James Prescott Joule, who tackled the problem of determining exactly how much heat was produced by a given amount of work. If energy were truly conserved, once all its forms were taken into account, then the same amount of heat should be developed from a given amount of work, regardless of the form of energy being used to produce the work.

Joule began his measurements in 1840 and continued them for a generation. He produced heat by driving an electric current through a wire, by forcing water through thin glass tubes, by compressing and expanding gases, by churning first water, then oil, then mercury with paddle wheels, by friction on cast iron, by passing electricity through a solution, by rotating a coil of wire between the poles of a magnet.

In each case he calculated the amount of work that had been done and the amount of heat that had been produced and found that the same amount of work, of whatever

energy source, always yielded the same amount of heat. In this way he determined what is called "the mechanical equivalent of heat." It turns out that about 41,800,000 ergs of work are equal to 1 calorie of heat. It was in Joule's honor that 10,000,000 ergs were set equal to 1 joule, so we can say that 4.18 joules equal 1 calorie.

It seemed clear to Joule in the 1840s that it made sense to assume that energy was indeed exactly conserved in the real world, if all forms of energy were taken into consideration. At least he could never find a case where energy was definitely *not* conserved. This seemed clear, during that same decade, to two German physicists, Julius Robert von Mayer and Hermann Ludwig Ferdinand von Helmholtz.

It is Helmholtz who is today usually given the credit for being the first to state the principle of the "conservation of energy," formally and definitely, in a book he published in 1847. According to this view, energy may change its form but may neither be created nor destroyed. Another way of putting it is that there is as much energy in the universe now as there ever was, no more and no less; nor will there ever be a whit more or a jot less.

Mayer in his own writings made passing remarks that indicated he had accepted this principle as early as 1842. He also went further than Helmholtz in a manner that is of particular importance to the subject matter of this book. Mayer felt that the source of all the various forms of energy on earth was the radiation of light and heat from the sun. Furthermore, he felt that the energy exhibited by living things, and by man in particular, also had solar radiation as its ultimate source. The implication was that the law of conservation of energy applied to living, as well as to non-living, objects.

In suggesting this, Mayer was making a vital point. Mankind has always been aware of the gulf that separated life from non-life and, in fact, I began this book with a consideration of that gulf. The rules that were developed by scholars to describe the behavior of the universe had until the nineteenth century applied either to the living or the non-living, never to both. Planets moved in certain ordained and unalterable paths but living things moved at will. A stone fell and smoke rose, but a human being could either fall or rise as he chose. On the other hand, a man or mouse might feel fear or might exhibit anger, but a brick could do neither.

Now, for the first time, there was the thought that a law of nature, the law of conservation of energy, might apply to living and non-living alike. This was a crucial step toward the understanding of life; perhaps the first to lift the study of living tissue above the merely descriptive.

Yet this would seem to carry us away from the purpose of this book as I described it in the first chapter. There, I suggested that a consideration of work and energy would lead to a better distinction between life and non-life than that which we arrived at by intuition only. And here, we seem to be moving in a direction that, far from helping us in this respect, appears to be blurring the distinction altogether.

But don't give up! More is involved in nineteenth-century studies of energy than mere conservation, as we shall eventually see.

Chapter 5

THE FLOW OF HEAT

The acceptance of the fact that energy is conserved meant, in some ways, a fundamental change in the human view of the universe, a change that has not entirely permeated human thinking even yet. In all ages men have believed that faith could move mountains; that a fairy godmother's wand or Aladdin's lamp could build a castle in the twinkling of an eye; that seven-league boots could whisk one through space . . .

Yet all this romance can be wiped out of existence by the question: "But where is the energy to come from?"

If we are to break down such lovely dreams by insisting on the reduction of everything to the double-entry accountancy of energy intake and outflow, let us make sure that we have a right to do so. In other words, how certain are we of the truth of the law of conservation of energy? Can it be proven that the law is true?

Well, the law cannot be proven true, at least not so far. No one has ever been able to suggest why energy should be neither created nor destroyed. We can only say that the most painstaking observations and the most ingenious experiments, from Joule's day to our own, have failed to provide us with a single clear case of energy being either created or destroyed.

As long as this is true, and as long as the assumption that the law is correct has proven limitlessly useful in a limitless number of ways, scientists will not abandon it. They cannot prove it but they are as certain of it as they can be of anything.

Only once, in fact, did the law of conservation of energy tremble to its roots. That was in 1896 when the phenomenon of radioactivity was discovered. It was then found that elements such as uranium produced energy continuously out of no apparent source. The problem grew crucial in 1898 with the discovery of new elements such as polonium and radium that radiated energy at millions of times the rate uranium did, and still with no discoverable source.

In 1905, however, the German-born physicist Albert Einstein published a new and very fundamental view of the universe (usually referred to as the "special theory of relativity") that rectified matters by calling mass to the rescue.

You see, mass itself also represents a conservation phenomenon. That this was so was not discovered till the late eighteenth century and was a surprise when it came. To casual observation, it would seem that a burning candle disappears as it burns and that its mass is destroyed. Again, if iron rusts it gains weight and this would seem an example of the creation of mass. However, such appearances are the result of not taking into account the conversion of solid or liquid matter into gases, or the absorption of gases from the atmosphere in the case of rusting.

In the 1770s, however, the French chemist Antoine Lavoisier began to apply quantitative measurements to chemical reactions in systematic fashion, with results that convinced other chemists of the importance of weighing and measuring. (Lavoisier earned the melodramatic title "Father of Chemistry" as a result.) By conducting combustion and rusting reactions in closed containers and making careful measurements of mass before and after, Lavoisier was able to show that when the masses of gases, either produced

or consumed, were taken into account, mass did not change perceptibly in the course of chemical reactions.

This introduced the "law of conservation of mass," which stated that mass could change its form but could be neither created nor destroyed. In other words, the total amount of mass in the universe was the same now as it had always been and as it would always be.

The laws of conservation of mass and of energy are very similar in appearance and throughout the latter half of the nineteenth century they were considered as twin generalizations: separate and distinct, but very much alike.

It was Einstein, in 1905, however, who was able to show, by deduction from a very few basic principles, that mass had to be considered a form of energy. In his view a very small quantity of mass was equal to a very large quantity of energy. His famous equation, $e = mc^2$, where e represents energy, m represents mass, and c represents the velocity of light, is an expression of that. From this equation it is possible to show that 1 gram of mass is equivalent to about 21,500,000,000,000 joules of energy.

The energy released in radioactive phenomena was therefore no longer mysterious. It was at the expense of a loss of mass so trifling that that loss could not be detected by the ordinary instruments available at the time. (Instruments capable of doing so have since been developed, and Einstein's theories checked out perfectly.)

The wholesale development of energy by nuclear bombs and reactors is again at the expense of mass. Even more important is the explanation now made available for the fact that the sun has poured out energy unceasingly throughout many ages. Helmholtz was concerned about this and advanced as an explanation the suggestion that the sun was slowly contracting and thus converting potential energy into heat and light. This had a logic to it, but involved certain inadmissible consequences with regard to the past history of the solar system and did not satisfy astronomers.

Once, however, it was learned that the hydrogen of the sun could be converted to helium with the loss of some mass and the consequent release of huge floods of energy, the existence of solar radiation (and that of stars, generally) could be explained within the law of conservation of energy.

Currently, it is understood that when we say "energy" we include "mass" as well. To make that perfectly clear,

however, we often speak of the "law of conservation of mass-energy."

Thus radioactivity, which seemed to shake the law of conservation of energy, ended by enthroning it more firmly and spectacularly than ever.

As a matter of fact, the principle had been intuitively accepted by practical men, long before physicists felt ready to accept it on the basis of experimental evidence. As early as 1775 the Paris Academy of Sciences refused to consider any device purporting to deliver more energy than it consumed. Such a device, if one were possible, could do work and still keep running indefinitely on the energy it was, in effect, creating. This would therefore be a "perpetual motion" device. The United States Patent Office today will not consider any perpetual motion device for patent until such time as the inventor brings in a working model—something that has never been successfully done.

Yet if the law of conservation of energy predicted that it was impossible to get more work out of any device than was equivalent to the energy put in, it seemed reasonable to hope that one could at least get out *as much* work as was equivalent to the energy put in. If one could not make a profit, so to speak, one ought at least to break even.

This possibility was a natural driving force behind theoretical investigations into the workings of the steam engine in the early nineteenth century. (The general public usually places too low a value on the theoretical advances that underlie the successful tinkering of the inventor. Watt himself, in his development of the steam engine, was guided by Black's researches on latent heat, which I described in the previous chapter. Yet how many people, who know of Watt's invention, realize that it was the underpinning of Black's theory that made it possible?)

Even Watt's best engines, useful and desirable though they were, did not possess efficiencies of better than 5 per cent. That is, the output of work was not more than 5 per cent as great as the input of heat energy. It seemed there was plenty of room to improve on this without violating any laws of conservation of energy. Even allowing for unavoidable losses through friction, heat conduction, and so on, there seemed no reason why efficiency couldn't be raised to some value approaching 100 per cent.

Or so it seemed until a French physicist, Nicolas Léonard

Sadi Carnot, began his studies on the flow of heat and its interrelationship with work, thus founding the science of "thermodynamics" (from Greek words meaning "the movement of heat").

This science grew in importance as the nineteenth century progressed, so that when the law of conservation of energy was clearly established, its fundamental importance to all matters involving energy-work interconversions gave it the alternate name of the "first law of thermodynamics." Sometimes it is referred to, with what almost amounts to reverence, as simply the "first law."

Carnot, who founded the science, had little time to do it in, for he died at thirty-six during a cholera epidemic in Paris. Nevertheless, he found time enough. In 1824 (at the age of twenty-eight), he published a small book entitled *Reflections on the Motive Power of Fire*, in which he described his researches and the reasoning that led him to decide on the extent to which heat could be converted to work.

He showed that only a certain fraction of heat energy could be converted to work, *and no more*, even under ideal conditions; that is, even if friction were to be eliminated and no energy were to be lost to the outside world. This fraction depended on the difference between the temperature of the steam and the temperature of the condensed water (using the absolute scale of temperature derived from Charles' law). If we call the first temperature T_2 and the second T_1 , then the fraction of heat that could, *at best*, be converted to work is:

$$\frac{T_2 - T_1}{T_2} .$$

The fraction of heat converted—or, in other words, the efficiency—is the same regardless of what substance is heated and cooled. It might be water, mercury, or anything else; only the temperatures involved count. For this reason, Carnot's equation applies to any "heat engine" rather than merely to a steam engine.

As an example let us suppose that the steam being used in a Watt engine is at a temperature of 100° C. or 373° K. while the water to which it is condensed is reduced in temperature to 5° C. or 278° K. The Carnot equation would reduce to:

$$\frac{373 - 278}{373} = 0.255.$$

Just about one quarter of the heat input can be converted to work and not a bit more, no matter how perfect and ideal the working of the engine. In order to increase the efficiency, one would have to use superheated steam to raise the value T_2 or make use of a liquid with both a higher boiling point and lower melting point than water, as for example mercury. Both dodges have been used but even modern steam engines do not surpass an efficiency of 0.25.

Carnot's equation not only pointed out the route by which the efficiency of heat engines could be increased, it also gave rise to a new generalization of the greatest importance.

If Carnot's equation is correct, then it implies that no work can be obtained from any engine where the hot and cold reservoirs are at the same temperature. In such a case T_2 would be equal to T_1 and could be substituted for it, so that Carnot's equation would become:

$$\frac{T_2 - T_2}{T_2} = \frac{0}{T_2} = 0.$$

To give a specific example, consider that there is energy in ocean water, even in ice-cold ocean water. If ocean water freezes, then each gram gives up 80 calories of heat (as Black showed when he first measured latent heat). Why, then, could not an ocean liner extract such heat and use it to run its motors? Some of the ocean water about the ship would be cooled down and might even freeze, but the volume of water in the oceans, and the heat it contains, is so vast that the percentage loss of heat required to run all the ocean liners in the world (and even all the machinery in the world) would be insignificantly small—and could be made up for by the heating effect of sunlight anyway.

No violation of the first law of thermodynamics is involved in this dazzling possibility. Energy is not being created out of nothing. Energy in one form (the heat content of the oceans) is merely being converted into another (the kinetic energy of the turning propellers), and conversion of energy

from one form to another is specifically allowed by the first law.

Nevertheless, no scheme of this sort has ever proved to work; no way of utilizing the heat energy of any substance that is all at one temperature has been devised. It is, in other words, a matter of the common experience of mankind that Carnot's equation holds under all conditions so far tested. In order to convert heat into work there must be a difference in temperature somewhere in the system.

Lord Kelvin expressed it this way: A transformation whose only final result is to transform into work heat extracted from a source which is at the same temperature throughout is impossible.

This is one way of stating what is now called the "second law of thermodynamics"—or, sometimes, just the "second law."

The second law is by no means as seemingly inevitable—once explained—as is the first. We might, perhaps regretfully, face the fact that energy cannot be created out of nothing (first law) but if the energy is there, how is it we can't use it all (second law)? If we can't make a profit, why can't we at least break even?

Well, we are all accustomed to considering the energy of motion since that is what we deal with every day, so perhaps the principle of the second law can be made more reasonable if we think of falling bodies.

A falling body can do work when it strikes the ground. The work may consist of killing a man or smashing a nut or just gouging out a hole, but work is done. Once the body has come to rest on the surface of the earth it can do no further work as long as it simply rests there. The work it did do was done by virtue of the fact that it moved from a point of high potential energy (say, the top of a mile-high cliff) to a point of lower potential energy (the foot of the cliff).

But the body at rest upon earth's surface still possesses energy. If a pit a mile deep is dug near it and it is dropped down that pit it can again be made to do work, as much work as it did in dropping from the clifftop. That is because a new difference in potential energy has been created.

So it is with heat. The heat in the ocean cannot be converted into work without either a reservoir colder than the

ocean or one that is warmer than the ocean. To put it as concisely as possible: just as it is not a rock, but a falling rock, that does work, so it is not heat but a flow of heat that does work.

Furthermore, as you know from your own experience, a falling body falls only one way—downward. If a body could be made to fall both ways, then the same object could be made to do work continually by first falling from the edge of the cliff down to the valley, then falling up to the cliff edge, then down to the valley, and so on. As a matter of fact, however, it only falls from a position of high potential energy to one of low potential energy.

The situation is analogous for all other forms of energy and, in particular, for heat. The German physicist Rudolf Julius Emanuel Clausius, in 1850, stated this explicitly by saying that in any spontaneous process (one that proceeds by itself without outside interference) heat always passes from the hotter body to the colder body and never vice versa.

Actually all of us know this, and we all take it for granted that a kettle of water over a burning gas jet will eventually boil and that warm water will melt ice. We find it natural to suppose that heat will flow from the hotter body to the colder one and not vice versa.

The same assumption is implicit in Carnot's equation, for if heat did not flow *only* from hotter to colder, then it could flow from one part of a one-temperature water reservoir to another part and work could be obtained in this way. Clausius' statement, therefore, is another way of stating the Carnot-Kelvin rule about the impossibility of getting work out of a one-temperature system and is, in effect, another way of stating the second law of thermodynamics.

Clausius not only made this statement about heat flow (which all of us know by experience and require no scientist to tell us) but he recognized this to be a vast generalization that played its part in all processes involving energy, under all circumstances and anywhere in the universe. For this reason he is generally credited with the "discovery" of the second law.

All this brings up another interesting point. Any heat engine must take a period of time to convert heat to work. None can conceivably do its work instantaneously. During

the time of working, heat is flowing from the hot reservoir out to the environment (which is usually cooler than the hot reservoir) and never vice versa. Moreover, heat is flowing from the environment into the cold reservoir (which is usually cooler than the environment) and never vice versa. Therefore, the hot reservoir is getting cooler and the cold reservoir is getting warmer. The difference in temperature is getting less with time and not as much energy is converted into work as would have seemed likely from the original temperatures of the hot and cold reservoirs.¹

It is possible to think of a heat engine in which both hot and cold reservoirs are at a higher temperature than the environment, but then the hot reservoir would lose heat more quickly than the cold. If both were at a lower temperature than the environment, the hot reservoir would gain heat more slowly than the cold. In either case, the temperature difference would decrease. In fact, in any system devised to turn any form of energy into work, there is a decrease in the energy difference with time, or else some of the potential work in the energy difference is not realized because of the necessity of overcoming friction or some other sort of resistance to the natural flow of energy.

Clausius, in 1865, invented a quantity that included this unavoidably lost energy, together with the absolute temperature, and called it the "entropy" of the system.

Now it is possible to envisage perfect heat engines, so shielded that heat neither enters from the outside environment nor is lost to the outside environment; an engine where nothing is lost through friction and so on. Under such conditions, all the work inherent in temperature differences is realized. There is no "lost energy." In any real process, however, there is friction, there is heat flow into and out of the environment. In general there is "lost energy" and the amount of "lost energy" increases with time (see Fig. 4).

Using Clausius' term, we can say, concisely: In any spontaneous process, entropy either does not change (under ideal cases) or it increases (in real cases).

¹In actual fact, the hot reservoir is not cooling and the cold reservoir is not warming because there is a fire constantly heating the steam and cold water constantly cooling the condenser. However, some of the heat of the fire is unavailable for work because it must be expended on maintaining steam temperature, and some of the heat lost in the cold water is consumed in maintaining the temperature of the condenser. So you see you are still not getting out all the work to which you would seem to be entitled by considering temperatures alone according to the Carnot equation.

Forgetting the ideal, we can just take it for granted that in the real world about us entropy always increases.

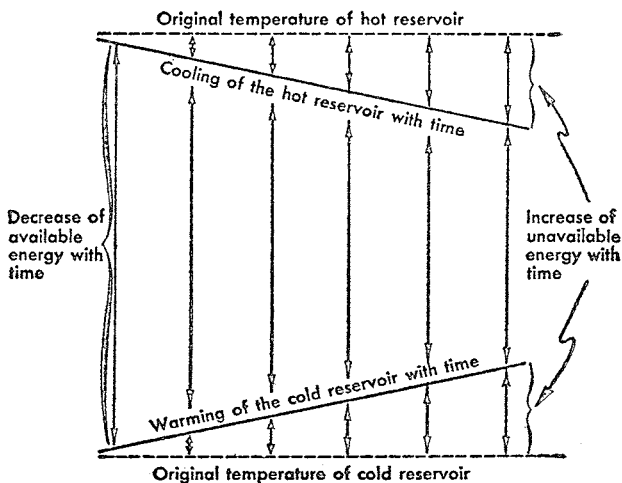


Figure 4. *Change of Available Energy with Time*

Since this is an inevitable consequence of the one-way flow of heat, this is still another way of stating the second law of thermodynamics.

To be sure, in paying reverence to the laws of thermodynamics, we must always remember that these laws apply to closed systems; that is, to portions of the universe that are imagined as being closed off from all influence by, or upon, other portions. If we deal with open systems, the laws can easily be made to appear to be violated.

For instance, suppose you watch a kettle of water and find the energy content to be continually rising until water begins to boil and then continues to boil until it is all gone. If you consider the kettle of water alone and nothing else, it would seem that energy has appeared out of nowhere in violation of the first law. This, however, you would not accept for a moment. You would feel certain, even without

looking, that under the kettle was a hot plate or a gas jet. Once these are included as integral parts of the system, you see, there is no longer a question of energy appearing out of nowhere, or any violation of the first law.

As other examples of apparent violations of the second law, where an open system is involved, consider that a body can move from ground level to the top of a tall building and that heat can flow from a cold spot to a hot one. You witness such actions constantly and if you concentrated on the moving body or the flowing heat alone you would have to conclude that the second law was violated.

But there is more to the system. The body moves from ground level to rooftop by means of an elevator powered by a motor. Heat flows from the cold interior of a refrigerator to the warm exterior by means of a motor's workings once again. If the motor is included as part of the system and the energy flow within it taken into account, you will find that the energy flow in the natural direction is greater than the flow in the unnatural direction, as evidenced by the moving body or flowing heat.

In other words, the local decrease in entropy (which invariably accompanies a flow of energy in the direction opposite to that indicated by the second law) in one part of the system is more than balanced by a larger increase of entropy in another part of the system. Taking the whole system into account, the *net* change is always in the direction of increasing entropy.

In fact, if we are not dealing with the imaginary constructions of the theoretical physicist but with reality itself, it becomes difficult to imagine a truly closed system. The outside environment always has some effect and must be included.

Thus the human body, if left to itself as part of a closed system, will die in a matter of minutes (for air must be excluded as part of the outside environment). From life to death might be considered the natural direction of change involving an increase in entropy. The fact that we do not die but persist in living, sometimes as long as a century or more, is only because our body does not make up a closed system in itself but is part of a larger system that includes the air we breathe and the food we eat. The interrelationships of all life forms and of all the inanimate background

of the planet on which we live make it reasonable to suppose that nothing less than all of earth can form a closed system in actual reality.

But even all of earth is not enough, for if we were to envisage the earth as a closed system, then life would still not last long, for the earth's surface would cool off, the oceans freeze, and so on. This, again, is the natural change and would involve an increase in entropy. That this does not happen is only the result of earth's being part of a larger system including the sun, the heat of which keeps the earth from cooling.

Even the solar system is not a truly closed system. It is affected by gravitational forces from the remainder of the galaxy and in various other subtle fashions. Every part of the universe has some effect on every other part and there is logic in supposing that there is no truly closed system other than that of the universe taken as a whole. It is only over the universe as a whole, in this view, that entropy must increase constantly. In any smaller portion, it is conceivable that entropy might locally and temporarily decrease, at the expense of a greater increase in the remainder of the universe.

Clearly, if the quantity of energy in the universe is finite and if entropy continues to increase, it must eventually reach a maximum where there is no longer any temperature difference at all. All the energy becomes unavailable; no process is left that can be spontaneous; no net changes can take place. This picture is that of the "heat-death of the universe," a point of view that grew to be popular in the latter part of the nineteenth century.

Now we can backtrack and see if it isn't possible to put into thermodynamic terms my original distinction between life and non-life. In the first chapter I said that life "made an effort" and that non-life did not.

You can now see that "making an effort" is invariably associated with a local decrease in entropy. If you are holding a rock in mid-air it is no effort to move it downward; you just let it go and it falls. It is an effort to move it upward, however.

The stone, left to itself, would move only downward, in the direction of "no effort." It cannot "make the effort" to move upward. Put it thermodynamically and you are saying

that the stone, left to itself, will undergo an increase in entropy but not a decrease.

A living organism can, however, "make the effort" of decreasing the entropy of at least part of the system in which it is involved (at the expense, admittedly, of increasing the entropy of the rest of the system still more). Even the simplest moving organism can enforce a local decrease in entropy by jumping, flying, climbing, walking, crawling, or swimming upward. Even sessile organisms such as an oyster or an oak tree can enforce local decreases in entropy in various fashions other than bodily movement.

Are we justified, then, in saying that the distinction between life and non-life is that life can bring about a local decrease in entropy and non-life cannot?

Certainly we are on the right track, but unfortunately the distinction as stated is insufficient. The heat of the sun brings about a local decrease in entropy when it evaporates water from the ocean and, in effect, raises huge masses of vapor high in the air. Forces within the earth raise huge mountains of rock miles high, again involving vast decreases of entropy. Yet neither the sun nor the earth is alive by any definition that modern man cares to use for life.

We must work at the distinction still further. Therefore, let us carry on.

Chapter 6

THE MOVING PARTICLES

Down to the very time of Joule, Kelvin, Mayer, Helmholtz, and Clausius, and the enunciation of the two great laws of thermodynamics, there was still no correct understanding of exactly what heat was; at least none that was accepted by the scientists working with heat.

This doesn't mean that the laws of thermodynamics are invalid, for they are based on observation and do not de-

pend on the nature of heat. However, it does mean that it was difficult to apply the laws of thermodynamics to energy devices other than heat engines, to devices where heat did not flow.

For instance, the human body (which is our chief concern in this book) certainly does work and yet it is at the same temperature (37° C.) throughout. Is this not a violation of the second law? We can be certain that it is not but in order to see *why* it is not we must consider the new conceptions of heat, and of matter itself, that were first put forward about 1800, although they were not completely accepted by the community of science until 1860.

To begin with, Galileo and Newton, as well as others, had speculated that heat might represent the motion of small subdivisions of matter. It was natural, perhaps, that they should think so, for their great researches had dealt with the energy of motion and by considering heat a form of motion they were fitting it into their great generalizations.

Through the eighteenth century, however, it became more and more popular to view heat as a weightless fluid, an "imponderable." This was a point of view first advanced by the Dutch physician Hermann Boerhaave about 1700. After all, there was no reason to think heat involved motion or vibration or that hot bodies vibrated or moved more rapidly than cold, because no such motion or vibration was to be seen. To say that small subdivisions of matter, too small to be seen, were vibrating too gently or too rapidly to be observed smacked of the mystical and was unpalatable to the scholars of the "Age of Reason."

To be sure, in 1738 a mathematics professor in St. Petersburg, Russia, named Daniel Bernoulli, had shown that, *if* gases consisted of tiny particles, the motion of those tiny particles would account for the pressure exerted by the gas. Furthermore, if the velocity of motion increased with rise in temperature, the changes in volume and pressure of the gas would change with temperature exactly as experiment showed that it would. However, at the time, this came under the heading of fanciful speculation. It seemed far better, to most, to turn to the imponderable-fluid theory, since so many other examples of these fluids were known (or were thought of as "known") at the time.

There was light, for instance, and electricity, and magnetism. There was even something called phlogiston, which was

thought to flow from a substance into air whenever that substance burned or rusted. Why should not heat be another of these? It consequently was accepted as such, and the heat fluid was named "caloric" from a Latin word meaning "heat."

At least one imponderable met its doom in the eighteenth century when, in the 1770s, Lavoisier demolished the phlogiston hypothesis and gave what we now consider to be the true explanation of combustion and rusting—not a loss of phlogiston but combination with oxygen of the air. This same Lavoisier, however, accepted caloric and lent it his prestige as foremost chemist of his time. That helped keep it alive for more than half a century in the face of mounting adverse evidence.

The first experiments that pointed strongly against the Lavoisier-encouraged imponderable-fluid theory and in favor of the heat-as-a-form-of-motion theory was carried out by an expatriate American, Benjamin Thompson. Thompson was a Tory who had fled America during the Revolutionary War. He was eventually given a title by the Elector of Bavaria in 1791 and, ironically enough, was married for a short time to the widow of none other than Lavoisier. It is by his title of Count Rumford that he is known to history.

In 1798, Count Rumford was supervising the boring of cannon for the Elector. As the boring tool ground its way into the metal cylinder to form the hollow interior of the cannon, a great deal of heat was evolved. To keep the heat under control, the bore was kept full of water that was replaced as it boiled away.

The proponents of the caloric theory understood that this sort of friction produced heat and they attributed it to the fact that it originally existed in the metal and as the borer cut into and ground up the metal the contained caloric was released. Rumford, however, noticed two things. First, there seemed an endless supply of caloric in the metal, for as long as he ground, so long was the heat produced. Secondly, when the borer grew dull so that it no longer cut into the metal very well, it did not stop producing heat; in fact, it produced *more* heat.

It seemed to him that the only reasonable interpretation was that the act of boring was producing the heat; that the work being done was converted into heat. (He even worked out a value for the mechanical equivalent of heat half a century before Joule, but the Count's value was far too high.)

He decided that work and heat were closely allied, and since the work he was observing involved mechanical motion, he decided that heat itself must be a form of motion.

A similar experiment was conducted the next year, 1799, by the English chemist Humphry Davy, who rubbed two pieces of ice together by a mechanical device, all held at a temperature somewhat below the freezing point of water. At that temperature, according to the old theory, there simply wasn't enough caloric in the ice for melting to take place. The mere act of rubbing with a mechanism that was also at a temperature below the freezing point of ice should not increase the amount of caloric present. Nevertheless, some of the ice melted under the rubbing. Davy, like Count Rumford, concluded that the energy of the motion was converted into heat, and that heat was a form of motion.

Despite this, the caloric theory persisted for decades.

The turning point came not in the study of heat itself but in the study of matter. The seventeenth and eighteenth centuries had seen a steady accumulation of information about the behavior of gases. I have mentioned Boyle's law and Charles' law in Chapter 4, for instance. That, and other information, cried out for some such interpretation of gaseous structure as Bernoulli had tried, unsuccessfully, to make.

Now, in 1803, an English chemist, John Dalton, went further. He suggested that not only gases but all matter was composed of small particles, far too small to see and themselves indivisible and indestructible. These he called "atoms," a word dating back to ancient Greek philosophy, for among the Greeks, some had advanced the same generalization. The Greeks, however, had advanced it as a pure speculation and it had not been accepted. John Dalton now based the generalization firmly on experimental observation—a crucial difference.

Even so, Dalton's atomic theory did not, by any means, take the world of chemistry by storm, largely because the idea of particles too small to be seen remained too hard to swallow. However, little by little the theory was found to explain more and more. It explained more than the behavior of gases alone (which Bernoulli had done); it explained a great deal about the chemical reactions of solids and liquids as well as gases. These reactions began to make remarkable sense if they were viewed as the interplay of different types

of atoms, associating themselves into groups (eventually called "molecules"), breaking up again, reassembling in different fashion, and so on. Though the atoms themselves remained too small to be seen, the increasing understanding of the behavior of matter entailed by the acceptance of atoms grew too large not to be seen.

The climax came in 1858, when the Italian chemist Stanislao Cannizzaro published a pamphlet detailing the manner in which the atomic theory could be used to explain chemical phenomena, and did it in so clear a fashion that no chemist could any longer avoid understanding. Attending a chemists' conference at Karlsruhe, Germany, two years later, Cannizzaro enforced his point of view by the power of both evidence and personality. Thereafter the atomic theory reigned in unquestioned supremacy. (It still does today although the picture of the atom now accepted is tremendously more complicated than that of the featureless, indivisible little spheres visualized by Dalton and Cannizzaro.)

With all matter, and gases in particular, pictured as collections of single atoms or, much more often, of the atom groups called molecules, there was a return to Bernoulli's early speculations as to the manner in which the motions of particles could explain gas behavior. New experiments were conducted and the results interpreted in the light of such a theory.

For instance, in the 1840s when Joule was turning work into heat in every way he could, he studied the expansion of a gas in a way that did *not* involve work. The gas did not lift a piston as it expanded or impart motion to anything. It merely expanded into a vacuum. When this was done, Joule could at first measure no rise or fall in temperature of the gas, as was to be expected of a process that involved no work.

By 1852, however, Joule, working along with Thomson (the later Lord Kelvin), had determined that a small fall in temperature does indeed take place under these circumstances. This is still called the "Joule-Thomson effect." The fall in temperature shows that the expansion of gas is not truly a workless process. Gas molecules do exert a very slight pull on each other and to move apart against that pull (motion against a resisting force) is work. The energy

for that work is drawn from the heat content of the gas and so the temperature drops.

This attractive force between molecules complicates calculations intended to describe the behavior of a gas, but theoretical chemists have simplified matters by dealing with an imaginary "perfect gas" in which the intermolecular attraction is supposed to be zero and in which the volume of the individual molecules is supposed to be zero also.

This creates an artificial situation, to be sure, but the resulting simplicity makes it possible for theoreticians to evolve simple generalizations. These generalizations apply approximately to the real world and, once evolved, can be adjusted to apply even more closely. This is clearly a more profitable line of procedure than it would be to attempt, from the start, a mathematical treatment of the actual world in all its complexity.

Thus Boyle's law and Charles' law apply exactly only to a perfect gas, and they can be combined to give a simple "equation of state" that describes the state of the gas in that it relates its temperature, pressure, and volume. Given any two of these properties, it becomes possible to calculate the third at once. In the case of real gases, the result is only approximate (though sometimes quite close). However, it is possible to modify the equation of state to allow for molecular volumes and intermolecular attractions of more than zero. A Dutch physicist, Johannes Diderik van der Waals, introduced one such modification (and quite a satisfactory one) in 1873.

This, I hope, explains and justifies the fact that during the 1850s various investigators (Clausius, for one) went on to elaborate and extend Bernoulli's work by assuming gases to be made up of zero-volume particles exerting no forces on each other, even though they knew this was not the actual situation as it existed. The result was a "kinetic theory of gases"; that is, one that dealt with particles in motion.

The theory was brought to a satisfactory pitch of excellence by the Scottish mathematician James Clerk Maxwell and the Austrian physicist Ludwig Boltzmann in 1859 and the years immediately following. The net result of their mathematical treatment (in the course of which they also assumed that gas molecules moved in random directions in straight lines, and collided with each other or with the walls

of the container under conditions of perfect elasticity, losing no energy in collision) was to relate molecular velocity to temperature.

Imagine two gas molecules colliding. They rebound, naturally, and in so doing change the direction of their motion and, very likely, their speed. The total momentum of the two particles must remain constant, however (see Chapter 3), and so certain limits can be set to the speeds they will possess after collision.

Even so, the redistribution of velocity after collision can take place in a variety of ways. If a quickly moving molecule strikes a slowly moving molecule, for instance, it is possible that the slowly moving one may come to a dead halt while the other streaks away more quickly than before with all the momentum possessed before the collision by both together. It is also possible, and much more likely, that the result is some intermediate speed for both, with the quickly moving molecule being slowed while the slowly moving molecule is sped.

(This same thing can be observed among billiard balls by those who play billiards. In Maxwell's time, atoms and molecules were, in fact, visualized as infra-tiny billiard balls. They are no longer so visualized; in fact, they are now known to be so complex in structure that it is impossible to visualize them with reasonable accuracy as anything at all with which we are familiar. Nevertheless, for the purposes of the kinetic theory of gases, the billiard-ball analogy remains adequate.)

It is impossible to say just how velocities will distribute themselves when a particular pair of molecules collide. However, over countless quintillions of collisions (as take place every second in any sample of gas large enough to be seen) it is possible to use statistical methods to determine the average distribution of velocities. By doing so, it can be shown that the molecules in any given sample of gas have a "most probable speed" about which the actual velocities of the molecules cluster. Some molecules, at any given time, move a bit more quickly than this most probable speed, while some move a bit more slowly. Others move quite a bit more quickly or quite a bit more slowly.

The rather complicated mathematical expression (called the "Maxwell-Boltzmann distribution law") (see Fig. 5) worked out by those two investigators made it possible to

calculate the percentage of molecules in a sample of gas that are moving in any particular velocity range (provided that

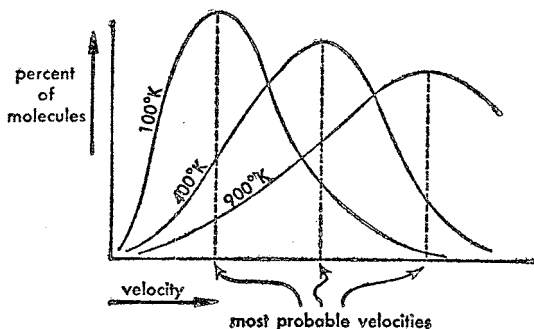


Figure 5. *Molecular Velocity and Temperature*

the temperature of the gas and the mass of the molecules composing it are known). If a graph is prepared from this equation relating velocity to the percentage of molecules, the curve bellies up to a peak at the most probable speed and declines thereafter.

The connection between temperature and the molecular motions within a gas can now be stated simply. In any gas, at any temperature, individual molecules may, at any instant of time, have a velocity varying from zero to very high values. However, the most probable velocity turns out to be proportional to the square root of the absolute temperature. As the temperature rises, the most probable velocity of the gas molecules increases. If the absolute temperature increases fourfold, the most probable velocity doubles.

For any particular gas, then, temperature is related to the most probable velocity of the molecules. If we are to generalize for all gases, however, the mass of the molecules must be taken into account. At any given temperature, according to the Maxwell-Boltzmann equations, the most probable velocity is inversely proportional to the square root of the mass. The more massive the molecules, the more slowly they move. The oxygen molecule is sixteen times as massive as the hydrogen molecule, so the hydrogen molecules move

four times (that is, the square root of sixteen) faster than do the oxygen molecules at any given temperature.

But kinetic energy, as I explained in Chapter 3, is equal to half the mass times the square of the velocity, and what I have already said implies that at any given temperature the average kinetic energy of the molecules of all gases is equal. The molecules of one gas may be more massive than those of another, but are also of lower velocity; if they are less massive, they are of higher velocity. In every case, though, the mass multiplied by the square of the velocity turns out to be equal. And if velocity goes up as the square root of the temperature, the kinetic energy, which depends on the square of the velocity, is directly proportional to temperature (see Fig. 5).

What we call "temperature," then, is the measure of the average kinetic energy of the molecules making up a gas.

As it turns out, this relationship holds for all substances and not merely for gases. To be sure, conditions are vastly different for non-gases. The molecules in a gas are far apart and move about independently. The molecules in liquids, however, are in contact and can merely slide over one another; while the molecules in solids are fixed in place and do not even slide. Nevertheless, the molecules of liquids slide with greater energy as temperature goes up, and the molecules of solids vibrate about their fixed positions with greater energy as temperature goes up.

All alike lose energy as the temperature drops and at absolute zero that energy is reduced to a minimum. (This minimum is usually considered to be zero, but modern theory indicates that even at absolute zero there is a very small quantity of energy remaining that cannot be removed. This has significance, for instance, in the odd behavior of liquid helium at temperatures near absolute zero.)

Now suppose that a sample of hot gas is brought into contact with a sample of cold gas. The kinetic energy of the molecules of the hot gas is higher on the average than that of the cold gas. In the collisions of a molecule of the one with the other the energy of the "hot" molecule is very likely to be greater than that of the "cold" molecule, even though there are some molecules in the hot gas that are moving more slowly than some molecules in the cold gas. (If this confuses you, consider that the standard of living in

the United States is higher than the standard of living in Egypt, so that an American chosen at random is very likely to be better off than an Egyptian chosen at random, even though some Egyptians are, nevertheless, wealthier than some Americans.)

As a result of the collision, then, the "hot" molecule may gain still more energy at the expense of the "cold" molecule and end by being "hotter" than ever. (I put "hot" and "cold" in quotation marks, because these concepts do not really apply to individual molecules but only to systems containing large numbers of molecules.) However, the greater likelihood, as I said earlier, is that the "hot" molecule will lose energy and the "cold" one gain so that both will end at some intermediate level of energy, and therefore of temperature.

If a large number of collisions are considered, then only a very few will take place in such a fashion that the difference in energy will increase while the vast majority will take place in such a fashion that the difference in energy will decrease. Unlikely as it is that a single "hot" molecule will gain further energy as a result of a collision with a single "cold" molecule, it is far, far more unlikely that a large number of "hot" molecules will all, simultaneously, do so at the expense of a large number of "cold" molecules.

As a consequence, by sheer random collisions, the average kinetic energy of the molecules, and therefore the temperature, of a hot gas (or any hot object) will fall when placed in contact with cold gas (or any cold object) while that of the cold gas (or any cold object) will rise. Heat will flow from the hot object to the cold object until the average kinetic energy of all the molecules involved reaches some intermediate value.

How does this differ from the older imponderable-fluid theory of heat? Well, in the older theory, heat flowed from high temperature to low in obedience to a second law of thermodynamics that allowed neither exception nor modification. According to the kinetic theory, however, the flow of heat was the statistical result of random collision and could, conceivably, be reversed, so that the second law was no longer to be considered absolute. It is possible, after all, that so many "hot" molecules will just happen to gain energy from "cold" molecules that heat will flow from the cold body to the hot body. Thus, a kettle of water set over

a fire might conceivably, by kinetic theory, freeze while the fire grew hotter.

To be sure, the chances of this happening (as worked out from Maxwell's equations) are so unimaginably small that, if the entire known universe consisted of nothing but kettles over fires, the probability that even a single one of them would freeze during the entire known duration of the universe would be so small that we could not reasonably hope to witness any examples of this odd phenomenon.

The difference between an invariable flow and a flow that deviates immeasurably from being invariable is all-important philosophically but seems so insignificant in any practical sense that Maxwell felt the need to dramatize it somehow. He therefore pictured a tiny being at a stopcock between two containers of gas, that were at equal temperatures to begin with. This being behaved in the following fashion: When a fast-moving molecule in the right-hand container approached he let it through the stopcock, while rejecting all the slow-moving molecules. On the other hand, when a slow-moving molecule in the left-hand container approached he let that through the stopcock, while rejecting all the fast-moving molecules. In this way the left-hand container would accumulate fast-moving molecules and grow warmer, while the right-hand container would accumulate slow-moving molecules and grow cooler. Heat would continue to flow from the cold body to the hot body, a situation inconceivable by the old theory but conceivable by the new one. "Maxwell's demon," as the imaginary being was called, has been famous ever since as the ideal entropy-reverser.

And yet, you know, we need not search for demons only in fantasy. Something like Maxwell's demon exists in real life. After all, statistics average out the behavior of molecules in so nearly invariable a fashion only because under ordinary circumstances we deal with such vast numbers of molecules. As the number of molecules in a system decreases, the chance of noticeable deviation from the statistical average increases. For very small numbers indeed, smooth uniformity breaks down altogether.

An example of what I mean arose out of an observation made back in 1827 by a Scottish botanist named Robert Brown, who observed that pollen grains suspended in water had an odd wavering movement in this direction and that, apparently at random. At first he thought this might be the

result of the latent life present in the pollen grains but it was quickly found that any small object (a particle of dye, for instance) underwent the same random movement. This has been called "Brownian movement" in honor of its discoverer ever since.

About 1890, Boltzmann showed how this could be explained by kinetic theory. He pointed out that any object suspended in water was bombarded on all sides by the molecules of water surrounding it. Statistically, the number of molecules striking it in any particular direction would be equal to those striking it in any other and the net effect would be that the object would be motionless.

Of course the numbers wouldn't be exactly equal but under ordinary circumstances the unevenness would be too small to be perceived. If a quintillion molecules hit in one direction and a quintillion and twelve in the other, the lack of balance would be unnoticeable. As the suspended body was made smaller, the total number of molecules striking it per second would be smaller also and the same small unevenness would become more nearly noticeable. By the time the body was small enough so that only one hundred molecules were striking it on one side and one hundred and twelve on the other, those same twelve unbalanced strikes that meant nothing in a quintillion would now mean a great deal. Boltzmann showed that if the particle were small enough the random motion of molecules, with a slight overbalance first in this direction, then in that, would account for the Brownian movement.

Here we have a kind of Maxwell's demon moving a particle that, under the old theory, ought to have remained motionless. Of course the movements are completely random and, in the long run and over many particles, get nowhere, so it is a drunken Maxwell's demon we are dealing with. (And, in fact, such a motion where no step has any necessary relationship with the previous one is sometimes called a "drunkard's walk.")

In 1905, Albert Einstein worked out a thorough mathematical treatment of the Brownian movement and delicate experiments in the 1920s and 1930s actually measured the velocities of molecules in gases and found they were distributed in agreement with the predictions of the kinetic theory.

It is possible to view entropy, too, in the light of kinetic theory. In order to do so, let us consider what we mean by the words "order" and "disorder."

Intuitively, we consider something "in order" that is methodically arranged according to some logical system. Something that is "in disorder" is higgledy-piggledy, with no discernible system to the arrangement.

Another way of looking at it is to say that something that is "in order" is so arranged that one part can be distinguished from another part; while something is "in disorder" when no part can be distinguished from any other. The clearer the distinction from part to part, the "more orderly" the system is; the fuzzier the distinction, the "less orderly" it is.

One commonly used example of order is that of a pack of cards arranged by suits and in order of value within each suit. This is perfect order, for you can tell each smallest portion of the deck from any other. You know that the first card is the ace of spades, the second card the deuce of spades, and so on through hearts, clubs, and diamonds. Given any card, you know its position in the deck; name any position and you know the particular card involved.

If the deck, on the other hand, is arranged in suits, but with no particular order within the suits, then you are less well off. If you are shown a spade, you know you are somewhere in the first quarter of the deck, but you can't tell where in the first quarter. You can distinguish one part of the deck from another still, but only within relatively broad limits. The deck still shows some order but it is less orderly than the original completely arranged deck.

Again, if the deck is completely shuffled, you cannot from the sight of any card tell where in the deck you are. Nor, given a region of the deck, can you tell anything at all about the card that occupies it. You are now faced with "randomness" or with "complete disorder."

Now let's consider a condition where entropy is increasing, using for the purpose the case of heat flowing from a hot body to a cold. (This, indeed, is the classic case of increasing entropy.) As the heat flows, the hot body grows colder and the cold body warmer, until finally both reach some intermediate temperature.

But as this happens, not only entropy but also disorder is

increasing. To begin with, it is easy to distinguish the two bodies, even if otherwise identical in all ways, simply by the difference in temperature. As heat flows and the temperature difference between the two decreases, the distinction grows fuzzier, until at the end, when both reach some intermediate temperature, there is no temperature distinction at all. Disorder has increased along with entropy.

Again, if a sample of hydrogen and a sample of oxygen, both at the same temperature and pressure, are brought together, there will be a diffusion of one into the other until the whole volume is an even mixture of the two. Since this happens spontaneously, the mixing must involve an increase in entropy. Since the mixing also causes the original easy distinction between one part of the total gas volume and the other to vanish completely, there has thus also been an increase in disorder.

In fact it is possible to view an increase in entropy as involving an increase in disorder every single time. Entropy and disorder are analogous phenomena.

Since all spontaneous processes involve increases in entropy, all must therefore also involve increases in disorder. This is true when heat flows from hot to cold, or when different gases mix. You don't expect heat, left to itself, to flow from cold to hot, and you don't expect a mixture of gases, left to itself, to sort itself out into its separate constituents.

The same is true even of the homely example we have been using: the pack of cards. A deck of cards in perfect order will not disarrange itself if left to itself, simply because it undergoes no change at all. However, suppose you decide to place it under conditions where some change must occur but where the changes may be considered to be guided by chance alone. In other words, suppose you drop the deck of cards from the rooftop to the street below. The deck comes apart and flutters downward as single cards. If those cards are now picked up at random, you will expect to find the deck in complete disorder even though they started their fall in complete order. In this spontaneous process, both entropy and disorder have increased.

Shuffling is another example of a spontaneous process (in the sense that it moves the cards about randomly) that increases both entropy and disorder. In fact, honest card games depend upon the fact that honest shuffling produces disorder.

In order to restore a deck from disorder to order, you must sit down and perform the task a card at a time. (You must "make an effort," to use the phraseology of the first chapter, which is natural since you are decreasing entropy.) You would not expect, by throwing a disorderly deck into the air, or by shuffling it extensively, to find it suddenly come to complete order.

To be sure, it is possible that by a most unusual stroke of luck the shuffling could bring about a partial order or even a perfect one just as, very occasionally, a person is dealt a bridge hand consisting of thirteen spades. This is possible by the very same line of reasoning that makes it possible for a kettle of water to freeze when placed over a fire. The only reason that an individual has an actual chance to see a thirteen-spade bridge hand in his lifetime is that only fifty-two cards are involved. If there were as many cards in a deck as there are molecules in a kettle of water, his chances of drawing one quarter of them at random and finding them all of one suit would be about the same as having that kettle of water actually freeze.

Of all forms of energy, heat is the most disorderly. That is, substances that differ in heat content alone are harder to distinguish than substances that differ in the amount of light they are radiating or their position in a gravitational field, or electromagnetic field. Consequently, in all spontaneous processes, involving energy of types other than heat, some energy is always converted to heat, this in itself involving an increase in entropy.

By the same line of reasoning, under no conditions can all the heat in a system be converted to some form of energy other than heat (barring the kettle-freezing-over-a-fire type of chance) since that in itself would imply a decrease in entropy. Instead, if some of the heat is converted to another form of energy, the remaining heat undergoes an entropy increase more than enough to make up for the entropy decrease involved in the shift of energy from heat to anything else.

Thus we can see that the kinetic theory of gases evolved in the second half of the nineteenth century adequately explains the thermodynamic phenomena observed in the first half of the nineteenth century under the influence of the imponderable-fluid theory.

However, I also said at the beginning of the chapter that the new theory would offer a chance to explain how the human body could produce work when it was at a single temperature; something the old theory could not explain. This I have not forgotten and we are working methodically toward that. The key words for the purpose are "atom" and "molecule"—words that did not exist in the old theory but are central in the new.

Chapter 7

THE PARTICLES HOLD TOGETHER

It might be tempting to think of the human body (or of living tissue generally) as extracting work from a single-temperature system because some property of life enables it to reverse entropy. After all, I did make a beginning at defining life in terms of entropy decrease at the end of Chapter 5.

However, let us not focus too closely on heat as a source of work, just because the early thermodynamicists did. Kelvin's statement of the second law did say that heat could not be converted to work without a temperature difference, but there are forms of energy other than heat. The complete generalization of the statement would be that energy of any form could not be converted to work without a difference of intensity-level in that form of energy. A falling rock, after all, does work even though its temperature is precisely that of the cliff top it leaves and the cliff bottom it strikes. No temperature difference is involved, but a difference in potential energy is.

It remains then to discover the particular form of energy that is utilized by living organisms and out of which work is extracted. But the phenomenon of life is still too complex to tackle head on and we must cast about for something simpler that will serve as a stepping stone.

If we return to our heat engines, we could begin by ask-

ing where the energy comes from to create the hot reservoir in the first place. The answer, at once, is combustion. Wood, coal, oil, or gas burns and the energy so developed heats water to steam and makes the heat engine possible.

Well then, where does the energy of combustion come from? The wood, coal, oil, or gas, to begin with, is at the same temperature as the air with which it combines. To be sure, any fuel must be heated a bit before combustion starts, but the energy delivered by the combustion (which then continues on its own) is far, far greater than the initial heat of ignition. Where does all that energy come from?

This is a question of the greatest importance for us, because, as I shall explain later on, it was recognized, comparatively early, that there was something very similar between the burning of coal and the breathing of a human being. It would seem that if we can find the source of energy in, let us say, cold lumps of coal, we might be on the road to finding the source of energy of the human body.

In any case, let us try.

In the eighteenth-century heyday of imponderable fluids, coal was thought to be an example of a substance full of phlogiston. On burning, the phlogiston was liberated and it was that which created the heat and light associated with combustion. However, this theory led to untenable contradictions and Lavoisier smashed it.

A new explanation of the energy source was required and none offered itself until the atomic theory came along.

A clue to the proper answer was given in the previous chapter when I mentioned the Joule-Thomson effect. Gas expanding into a vacuum does work because the individual molecules in moving outward from the common center have to overcome the small attractive forces holding them together.

Thus there is a force of attraction between molecules just as there is a force of attraction between astronomical bodies; there are intermolecular chemical forces just as there are interplanetary gravitational forces. The existence of gravitational force gives rise to the concept of potential energy and the existence of intermolecular chemical forces gives rise, analogously, to the concept of "chemical energy."

To show how, let us begin by considering ice. It consists of water molecules in virtual contact and in orderly array.

(There is a variety of evidence for this that I won't go into. Such an orderly array of atoms or molecules makes up a "crystal" and most solids can be made to exist in crystalline form.)

At a given temperature below 0° C. the individual molecules of ice vibrate to and fro about the ideal position in this array. The kinetic energy of this vibration is insufficient, however, to overcome the intermolecular attractive force that keeps the molecules put, so that no molecule leaves its position permanently.

If the ice is heated, however, its temperature rises and the molecules vibrate more quickly and possess more and more kinetic energy. When a temperature of 0° C. is reached, the kinetic energy of the molecule becomes large enough to overcome the intermolecular attraction and the ice crystal begins to break down. Individual molecules begin to move freely and are no longer bound into a given position with relation to other molecules. In short, solid ice begins melting to form liquid water.

As heat is poured into melting ice, none goes into increasing the kinetic energy of the individual molecules, so that the temperature does *not* go up. Instead, all of the heat is applied in overcoming the force of intermolecular attraction so that more and more of the ice melts. The water molecules in the melt do not possess more kinetic energy than the water molecules bound in the crystal, so that the liquid water being formed is at the same temperature as the ice that is still unmelted. Nevertheless, the water must contain more energy in some form than does the ice, for the molecules in the water have overcome the binding forces holding them in crystalline array, while the molecules in the ice have not. This additional energy, not in the form of heat, is the chemical energy I mentioned.

Liquid water at 0° C., then, contains more chemical energy than does ice at 0° C. While ice is melting, the added heat is being converted into chemical energy instead of kinetic energy so that the ice melts without a rise in temperature. It is only after all the ice has melted that the heat being added can be converted to kinetic energy and applied to the raising of temperature again.

A gram of ice at 0° C. absorbs eighty calories in the process of conversion to a gram of water at 0° C., so liquid water possesses eighty calories per gram more chemical en-

ergy than ice at that temperature. This is that latent heat of fusion discovered by Black in the days of the caloric theory. He was unable to explain the phenomenon successfully (though he made a heroic try) since he used only imponderable fluids as theoretical background. We are more fortunate.

But we are not through. In liquid water (or any liquid) the constituent molecules may not be fixed in any particular position, but they do remain in contact, sliding over one another. There are still forces that keep the molecules from assuming complete independence. The addition of further heat results in two changes. Some of the heat is converted into kinetic energy so that the molecules vibrate more quickly and the temperature goes up. Some of the heat is converted into chemical energy, enabling a molecule to break away from the rest without rise in temperature. Thus water vapor is formed.

At any temperature, liquid water (or any liquid) is always vaporizing to some extent. If the liquid is in a sealed container, some of the vapor is always condensing back to liquid and an equilibrium is established so that at any given temperature there is a certain "vapor pressure." The higher the temperature, the greater the pressure.

Under ordinary conditions, atmospheric pressure is greater than the vapor pressure of water, so that the outside-in push of the atmosphere, so to speak, more than counters the inside-out push of water vaporization. At 100° C., however, the vapor pressure rises to equality with the atmospheric pressure and water pushes aside the restraint of the atmosphere and vaporizes rapidly enough to produce what we call boiling. (Naturally, if air pressure is lowered, vapor pressure rises to equality at a lower temperature. Water boils at a lower temperature at high altitudes for that reason and making hard-boiled eggs becomes difficult.)

While water is boiling, added heat is converted into chemical energy, changing liquid water at 100° C. into steam at 100° C. Again one gram of steam at 100° C. has 540 calories more chemical energy than one gram of water at 100° C. This is the latent heat of vaporization.

Naturally when steam condenses back to water, it must give up this chemical energy again, which is converted into heat. Since a gram of condensing steam gives up 540 calories, while a gram of water cooling from 100° C. to 0° C.

gives up only 100 calories, you can see that most of the work obtained from a steam engine is extracted from the condensing steam rather than from the subsequent cooling of the water.

Water vapor at any temperature contains more chemical energy than liquid water at the same temperature. It is for this reason that perspiration acts to cool the body. The chemical energy required to vaporize perspiration is drawn from the heat content of the body with which the perspiration is in contact, so that the body is cooled. When the humidity is high so that recondensation almost balances evaporation, perspiration collects in droplets, the cooling effect diminishes, and we go about saying: "It's not the heat, it's the humidity."

Even this is not the end, for the water molecule itself is not a simple particle but is made up of three atoms, two of hydrogen and one of oxygen. There are forces holding these atoms together within the molecule that are far stronger than those holding neighboring molecules together.

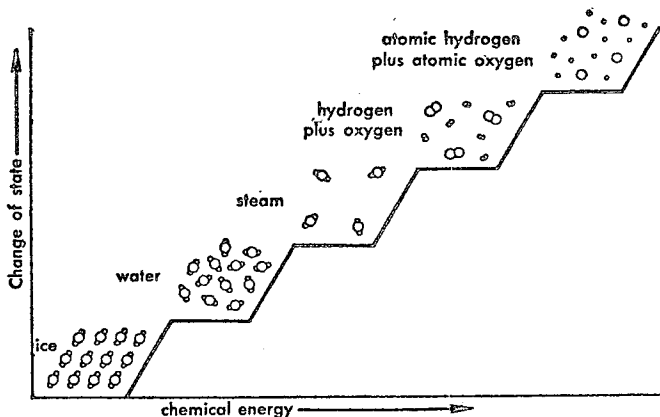


Figure 6. *Chemical Energy and the Disassociation of the Water Molecule*

These interatomic forces are so much stronger than the intermolecular forces that an analogy with ropes and cables

seems almost inevitable. The interatomic attractive forces are usually referred to, therefore, as "chemical bonds."

As heat is added to water (see Fig. 6), not only do molecules gain kinetic energy and speed their vibrations, but the atoms making up the molecule do so as well. However, even when enough heat is added to enable the water molecules to break all intermolecular attractions, the atoms within the molecule still lack the energy to break the far stronger interatomic attractions that hold the molecule together. Even when steam is heated quite strongly, the water molecules making it up remain intact.

However, if the temperature is raised to several thousand degrees, the energy of vibration of the atoms begins to be great enough to break the chemical bonds and water begins to dissociate into hydrogen and oxygen. At 3000°C . about one quarter of the water molecules have broken up into hydrogen and oxygen. A mixture of hydrogen and oxygen, then, contains more chemical energy than does the same mass of water at the same temperature.

The hydrogen and oxygen formed from the decomposing water are still not simple particles. Each is composed of a pair of atoms. Two hydrogen atoms are bonded to form a hydrogen molecule and two oxygen atoms are bonded to form an oxygen molecule. The bonds between these atom pairs can also be broken if enough heat is funneled into the substance. At a temperature of 3500°C . about a quarter of the oxygen molecules are broken up into single oxygen atoms ("atomic oxygen") while at that same temperature more than a third of the hydrogen molecules are broken up into atomic hydrogen. Oxygen and hydrogen in atomic form contain, again, more chemical energy than does the same mass of oxygen and hydrogen molecules at the same temperature.

Higher temperatures still begin to decompose the oxygen and hydrogen atoms themselves into still smaller particles, but we will not carry the story past the point of intact atoms at this time.

From what I have said, you can see that interatomic and intermolecular attractive forces are of varying strength and that the disruption of a particular bond, its formation, or, for that matter, the substitution of one type of bond for another, involves the gain or loss of an amount of chemi-

cal energy that is different in each specific case. When chemical energy is gained, energy in other forms (usually heat) must be consumed, to meet the requirements of the first law. Contrariwise, when chemical energy is lost, energy in other forms (usually heat, often light as well) must come into being.

Now we can answer the question as to where the energy of burning coal comes from. The heat and light associated with combustion arise out of the loss of chemical energy associated with the changes in the nature of the chemical bonds in the coal and in the oxygen with which it combines.

To go into further detail concerning the types of interactions involving changes in the nature of chemical bonds (i.e., "chemical reactions") that will deliver energy in the form of heat and light, we shall have to make use of a kind of chemical shorthand. This is not to make matters more esoteric but, on the contrary, to make them easier to understand. Fortunately, chemical shorthand is very straightforward and not at all difficult to grasp.

To begin with, there are presently known 103 different chemical elements, by which I mean 103 different varieties of atoms, each of which, in conglomeration, forms substances with distinct and characteristic properties. For the purposes of this book, it is not necessary to be familiar with all 103. A dozen or so of the most important will suffice.

Each element has, of course, its own name, which may be as simple as "tin" or as complicated as "praseodymium." In addition to the name, however, chemists have agreed, internationally, on a symbol for each. The symbol is in the form of one letter (the initial of the Latin name, which is usually, but not invariably, very similar to the English name) or of two (the initial plus a letter from the body of that name).

In addition, early investigations into the nature of the atoms revealed that they differed principally in their masses. That is, the individual carbon atom is twelve times as massive as the individual hydrogen atom, while the individual magnesium atom is twice as massive as the individual carbon atom, and so on. The atoms of the various elements are therefore assigned numbers representing their relative masses. The mass of the oxygen atom was arbitrarily set

at 16.0000 and all other atoms are given "atomic weights" in due proportion.¹

In Table 1 are listed some of the elements we'll be dealing with in this book, along with the symbol and the atomic weight for each.

As you can see, the only symbols among these elements that aren't perfectly obvious are those for sodium, potassium, and iron. In these cases, the Latin names are different from the English ones, being, respectively, *natrium*, *kalium*, and *ferrum*. It is annoying to have to memorize abbreviations that seem to have no relation to the name itself but it can be done. It can even become second nature, as in the abbreviation "lb." for "pound," where "lb." actually stands for the Latin word for "pound," *libra*.

The atomic weights, as given, are the most precise available (by the new 1961 standards) for the record. However, there is no need to try to remember all the decimal places. Many values are quite close to whole numbers and it is

Table 1

SYMBOLS AND ATOMIC WEIGHTS

Element	Symbol	Atomic Weight
Hydrogen	H	1.00797
Carbon	C	12.01115
Nitrogen	N	14.0067
Oxygen	O	15.9994
Sodium	Na	22.9898
Magnesium	Mg	24.312
Phosphorus	P	30.9738
Sulfur	S	32.064
Chlorine	Cl	35.453
Potassium	K	39.102
Iron	Fe	55.847
Iodine	I	126.9044

usually sufficient to consider the atomic weight of hydrogen as 1, that of carbon as 12, that of sodium as 23, that of

¹ In 1961, a very small change was made in order to adjust atomic weights in accordance with modern methods of measuring atomic properties. One variety of carbon atom is set at 12.00000. Compared with that, the atomic weight of oxygen is now set at 15.9994.

sulfur as 32, and so on. About the only cases where fractions might reasonably be remembered would be that of chlorine as $35\frac{1}{2}$ and magnesium as $24\frac{1}{3}$.

There is a definite saving in time and effort in writing C in place of "carbon" or H in place of "hydrogen," particularly for chemists who spend much of their writing time on the names of the elements. The saving becomes enormous when it is necessary to deal with molecules made up of combinations of atoms. Here, the use of symbols is not only a matter of timesaving, but one of a vast advance in clarity; a fact that will be borne in upon you during the course of this book.

As a small case in point right now, consider oxygen. It occurs, under normal conditions, as a gas consisting of molecules, each of which is in turn made up of a pair of oxygen atoms. Consequently, if we simply say "oxygen" it is not clear whether we mean the atom or the molecule. We could distinguish by saying "atomic oxygen" or "molecular oxygen," but not only does this involve polysyllabics, we are not specifying the exact constitution of the oxygen molecule. There is a form of oxygen molecule (usually called "ozone") that is made up not of two but of three oxygen atoms.

To simplify matters, it is only necessary to combine small numerals with the symbol for the element and we have a "chemical formula" that tells us the structure of the molecule. Thus, the symbol O stands for atomic oxygen. (One might write it as O_1 , specifying that the individual particles making up the substance consist of single oxygen atoms, but the numeral "1" is always understood when a symbol is written by itself.)

On the other hand, molecular oxygen of the ordinary type is written O_2 , while ozone is written O_3 .

Similarly, nitrogen, hydrogen, and chlorine are all gases made up of molecules containing pairs of atoms and are written N_2 , H_2 , and Cl_2 when the ordinary gas is referred to.

(Sometimes a substance is not made up of clearly defined molecules. A lump of carbon, such as a piece of coal, for instance, contains a vast number of carbon atoms that are held together by chemical bonds in a complicated manner. The entire piece of coal may even be viewed as one gigantic molecule. However, to simplify matters, elements in their solid state are usually symbolized simply as atoms, the actual

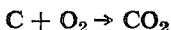
interrelationships of those atoms being ignored. Carbon in its solid, elementary form is therefore symbolized simply as C.)

This system of symbols and numbers can be used when molecules made up of more than one variety of atom are being dealt with. (Substances with molecules containing atoms of more than one element are referred to as "compounds.") Thus, the compound, hydrogen chloride, is a gas composed of molecules that are each made up of two atoms, one of hydrogen and one of chlorine. The formula is HCl.

In the case of water, which has molecules made up of two hydrogen atoms and an oxygen atom, the formula is H_2O (the one formula all non-chemists know), while that of carbon dioxide, with one carbon atom and two oxygen atoms in the molecule, is CO_2 .

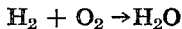
Each molecule has a "molecular weight" that is merely the sum of the atomic weights of the atoms that compose it. The oxygen molecule, O_2 , has a molecular weight equal to 16 plus 16, or 32. That of carbon dioxide, CO_2 , is 12 plus 16 plus 16, or 44, and so on.

Formulas can be used to express chemical reactions far more briefly and clearly than words can. For instance coal, which is almost pure carbon, will, when it burns, combine with the oxygen in the air to form carbon dioxide. This chemical reaction is written:



which can be read "carbon plus oxygen yields carbon dioxide" and is an example of a "chemical equation."

Similarly, hydrogen will combine with the oxygen in the air to form water. One way of expressing this might be:



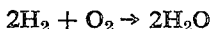
but chemists would not be satisfied with that. I shall explain why.

In the first reaction above, the one involving the combination of carbon and oxygen, you will see that there are no atoms left unaccounted for in moving from one side of the equation to the other. You start with one atom of carbon (representing the coal) and two atoms of oxygen (represent-

ing the oxygen molecule) and you end with the molecule of carbon dioxide, which contains the one atom of carbon and two atoms of oxygen with which you started. This is a "balanced chemical equation."

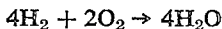
The second equation, the one involving the combination of hydrogen and oxygen, is a different matter. There you start with two atoms of hydrogen (representing the hydrogen molecule) and two atoms of oxygen (representing the oxygen molecule) but end with a molecule of water containing both hydrogen atoms, to be sure, but only one of two oxygen atoms. This is an "unbalanced chemical equation" and is misleading because it implies that one atom of oxygen has disappeared, which cannot, in actual fact, happen. Furthermore, equations can be used as guides for numerical calculations, but only balanced equations can be used as accurate guides (see Fig. 7).

To balance an equation, different quantities of starting materials may have to be indicated. Thus, a balanced equation for the combination of hydrogen and oxygen would be:

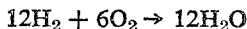


in which case you begin with a total of four hydrogen atoms and two oxygen atoms and end with a total of four hydrogen atoms and two oxygen atoms. Balance!

You might also, as far as sheer arithmetic is concerned, write:



or



and achieve balance, but it is customary (for simplicity's sake, if nothing else) to use the smallest whole number of molecules that can be managed.

At least it is the usual rule to use whole numbers only, but it is not the invariable rule. When the energy developed in a chemical reaction is being considered (which is just what I am going to do eventually) it is often customary to write " $\frac{1}{2}\text{O}_2$ " to indicate a single oxygen atom. This is *not* good practice ordinarily because it implies that "half mole-

cules" exist under ordinary circumstances, and this is *not* so. Still, this is a useful device in making calculations involving

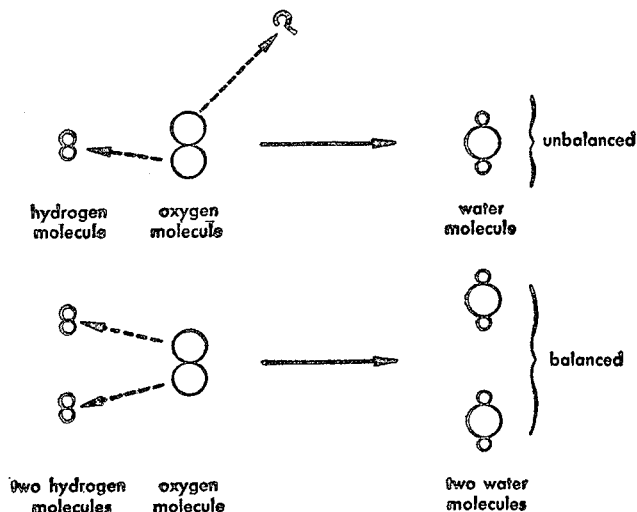
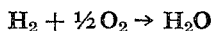


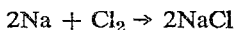
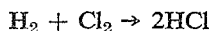
Figure 7. *Balanced Chemical Equations*

chemical energy, so the equation for the combination of hydrogen and oxygen can be written:



This equation, you see, begins with two hydrogen atoms and one oxygen atom (half a molecule) and ends with two hydrogen atoms and one oxygen atom, so that it is balanced.

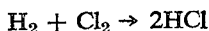
As examples of equations not involving oxygen, there is the union of hydrogen and chlorine to form hydrogen chloride, and of sodium and chlorine to form sodium chloride (or common table salt). The equations involved are these:



You can see for yourself that this is the simplest way (using

whole numbers only) of writing the equations and having them balanced.

Once the course of chemical reactions is made clear and simple by the use of symbols in the form of balanced equations, it becomes possible, as I said earlier, to use those equations in making various calculations. Consider, for instance, the equation representing the combination of hydrogen and chlorine:



This indicates that, in the course of the reaction, one molecule of hydrogen reacts with one molecule of chlorine to form two molecules of hydrogen chloride. Equal numbers of molecules of hydrogen and chlorine react—one with one, two with two, ten with ten, a billion with a billion. However, this does not mean that equal weights of the two gases combine, for one molecule of chlorine is much heavier than one molecule of hydrogen; and the weight of chlorine involved in the reaction is correspondingly heavier than the weight of hydrogen involved.

Since the atomic weight of hydrogen is 1 and that of chlorine is $35\frac{1}{2}$, the molecular weight of H_2 is 1 plus 1, or 2, and the molecular weight of Cl_2 is $35\frac{1}{2}$ plus $35\frac{1}{2}$, or 71. Whether one molecule of hydrogen combines with one molecule of chlorine, or two molecules of hydrogen with two molecules of chlorine, or ten with ten or a billion with a billion, the ratio of the masses involved is always 2 to 71. Thus, 2 ounces of hydrogen will react with 71 ounces of chlorine; 2 pounds of hydrogen will react with 71 pounds of chlorine; 2 tons of hydrogen will react with 71 tons of chlorine, and so on.

Chemists, who are wedded to the metric system, will most frequently consider that 2 grams of hydrogen will combine with 71 grams of chlorine.

(Chemists learned that 2 grams of hydrogen combined with 71 grams of chlorine, by actual measurements, before they knew the detail of the atomic make-up of the molecules. It was only after they had deduced, from various observations, that the combination was a one-to-one affair that they were able to conclude that the chlorine molecule was $35\frac{1}{2}$ times as massive as the hydrogen molecule, and that

the chlorine atom was $35\frac{1}{2}$ times as massive as the hydrogen atom. It was from the proportion by weight in which various elements combined, therefore, that it was first possible to work out the atomic weights of the various elements.)

Now it so happens that the number of hydrogen molecules in 2 grams of hydrogen is about equal to 602,000,000,000,000,000,000,000. This large number is called "Avogadro's number" because its use stems back to researches conducted, and conclusions reached, by the Italian chemist Amedeo Avogadro about 1810. Avogadro's number is usually symbolized as N and, since it is a good deal easier to write that than it is to write the number, I shall use N exclusively hereafter.

We can say, then, that N molecules of H_2 weigh 2 grams. Since the individual Cl_2 molecule is $35\frac{1}{2}$ times as massive as the individual H_2 molecules, N molecules of Cl_2 must be $35\frac{1}{2}$ times as massive as N molecules of H_2 . Therefore N molecules of Cl_2 must weigh $2 \times 35\frac{1}{2}$ or 71 grams.

This same sort of reasoning can apply to any molecule. The weight of N molecules of any substance is to the weight of N molecules of hydrogen as the weight of one molecule of the substance is to the weight of one molecule of hydrogen.

The number, N , is chosen because the weight of N molecules of hydrogen is equal to the molecular weight of hydrogen expressed in grams. (That is, the molecular weight of hydrogen is 2 and the weight of N molecules of hydrogen is 2 grams.) The ratios expressed in the previous paragraph make it necessary, then, that N molecules of any substance must have a weight equal to the molecular weight of that substance expressed in grams. As already worked out, N molecules of Cl_2 weigh 71 grams and the molecular weight of chlorine is 71.

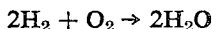
Since the molecular weight of H_2O is 1 plus 1 plus 16, or 18, we can say at once that N molecules of H_2O have a weight of 18 grams, while N molecules of CO_2 (with a molecular weight of 44) have a weight of 44 grams.

A weight equal to the molecular weight in grams is called the "gram molecular weight," which is customarily abbreviated as "mole." Thus, N molecules of any substance represent a mole of that substance. A mole of hydrogen weighs 2 grams, a mole of water weighs 18 grams, a mole of carbon dioxide weighs 44 grams, and so on. The weights are dif-

ferent but these quantities are characterized by the possession of equal numbers of molecules.

Naturally, where single atoms are being considered, the same line of argument holds. Thus, N atoms of carbon (C) weigh 12 grams, while N atoms of sulfur (S) weigh 32 grams. This, the *atomic* weight in grams, is properly called "gram atomic weight." For the sake of simplicity, I shall, however, let the abbreviation "mole" stand indiscriminately for "gram molecular weight" and "gram atomic weight."

The chemical symbols representing elements and compounds can be (and are) forced to do double duty. They can represent either one atom or molecule or they can represent N atoms or molecules. For instance, the equation:

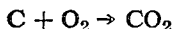


can be read "two molecules of hydrogen plus one molecule of oxygen yield two molecules of water." It can also be read "two moles of hydrogen plus one mole of oxygen yield two moles of water."

Since the molecular weights of hydrogen, oxygen, and water are 2, 32, and 18, respectively, one mole of each is (again) 2 grams, 32 grams, and 18 grams. The second way of reading the equation can therefore be translated into grams as follows: "4 grams of hydrogen plus 32 grams of oxygen yield 36 grams of water."

Matter, you see, is conserved. You begin with 36 grams of hydrogen plus oxygen and end with 36 grams of water. It is only a balanced equation that allows such conservation.

It is no trick at all, now, to see that the equation:



can be read in any of the following three ways:

(a) One atom of carbon plus one molecule of oxygen yields one molecule of carbon dioxide, or

(b) One mole of carbon plus one mole of oxygen yields one mole of carbon dioxide, or

(c) 12 grams of carbon plus 32 grams of oxygen yields 44 grams of carbon dioxide.

We now have the chemical language necessary to go on to a consideration of the measurement of energy changes involved in chemical reactions of the type which (we hope)

will prove analogous to those that take place in the human body.

Chapter 8

THE HEAT OF REACTION

There is no way of measuring chemical energy directly, but it is easy to measure energy in the form of heat. Suppose, then, that a particular reaction is allowed to take place in a closed container, after being set off, let us say, by a heated electric wire. The closed container would be surrounded by a fixed quantity of water that is kept well stirred. The whole system—reaction container plus water jacket—is insulated from the outside world in order to minimize loss of heat to, or gain of heat from, the environment.

Suppose it is a mixture of hydrogen and oxygen that has been ignited within the container. Combination takes place in a flash and water is produced. The chemical energy represented by the chemical bonds in the hydrogen and oxygen molecules is greater than that represented by the bonds in the water molecule. In passing from hydrogen plus oxygen to water, we therefore lose chemical energy. But this loss of chemical energy is exactly balanced by the gain of an equivalent quantity of energy in the form of heat.

As a result of the appearance of this heat within the container, the temperature of the water outside rises to a higher level. The stirring makes sure that this higher level is reached as quickly as possible and that all parts of the water are equally warm. The insulation makes it certain that any rise in water temperature is as a result of the heat developed within the container and not from heat leaking into the water from outside.

By measuring the amount of temperature rise in the known volume of water, the number of calories of heat delivered by the reaction ("heat of reaction") is determined, and this automatically measures the change in chemical ener-

gy that has taken place. To be sure, the energy evolved by the reaction may not be entirely in the form of heat. In ordinary combustion there is usually the formation of considerable light in addition to heat. Within the closed container, however, any light that is formed is quickly absorbed by the walls and converted to heat. (In the rather unlikely case that a reaction involves a gain in chemical energy, this must be balanced by the disappearance of an equivalent quantity of heat and a consequent fall in the temperature of the surrounding water.)

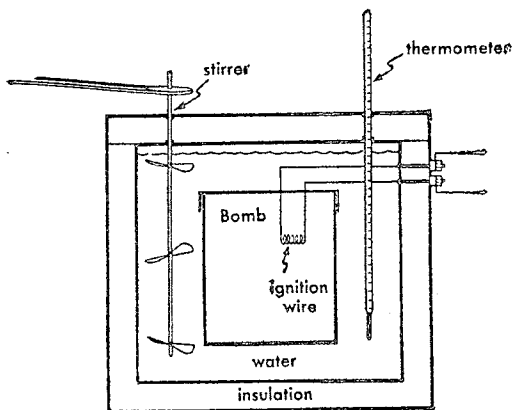


Figure 8. *A Bomb Calorimeter*

A container of this sort is a "calorimeter" ("heat measure" in Latin), and studies involving such devices are referred to as "calorimetry." Because reactions taking place within the chamber of a calorimeter are usually explosively fast, the chamber is referred to as the "bomb," and the whole device as a "bomb calorimeter" (see Fig. 8).

Modern calorimetric technique dates back to the experiments of the Danish chemist Julius Thomsen and the French chemist Pierre Eugène Marcelin Berthelot, who conducted hundreds of calorimetric determinations in the latter half of the nineteenth century.

By measuring the heat of reaction involved in converting a hydrogen/oxygen mixture to water, you do not determine the actual quantity of chemical energy present in the

hydrogen/oxygen mixture or in the water. You measure only the *change* in chemical energy in going from one to the other, the *difference* in the level of chemical energy between the hydrogen/oxygen mixture on the one hand and the water on the other. Actually, this change or difference is all the chemist is really interested in, for, by second law, any work that can be obtained out of a chemical reaction depends only upon the difference in levels of chemical energy, just as any work that can be obtained out of a heat engine depends upon the difference in levels of heat energy (i.e., temperature).

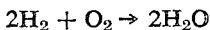
In physics and mathematics, a change in any quantity is often represented by the Greek letter "delta" (Δ) placed before the symbol for that quantity. Delta, which is the equivalent of our "D," stands for "difference." Thus, if we let energy be represented by the symbol E , then a change in energy would be ΔE , which would be read "delta E ." In the case of the reaction of hydrogen and oxygen to form water, ΔE could represent the chemical energy lost in the process or the heat energy gained. Both, of course, are equal in quantity.

Measurements of the heat of reaction can be made in two different ways, however. The substances making up the chemical reaction can be kept at constant volume, as they are, for instance, in calorimeters. The bomb of the calorimeter does not change volume in the course of the reaction. If the quantity of gas (and calorimetric experiments usually deal with gases or vapors) increases as a result of the reaction, the volume is nevertheless held the same, so the pressure must increase. If the quantity of gas decreases the volume still remains the same and the pressure decreases. The symbol, ΔE , refers specifically to the energy change under these conditions of constant volume and changing pressure. (Sometimes the symbol, ΔU , is used, but I shall stick to ΔE .)

On the other hand, it is possible to imagine the reaction as being conducted in a vessel, the top of which is a frictionless piston, that will move up and down freely but is airtight. In such a vessel pressure cannot change, for it must equal just the value required to balance the weight of the piston. If pressure increases, the piston moves up just far enough to allow the volume to increase and the pressure to decrease to the balancing point again. If the pressure drops, the

piston moves downward, decreasing the volume and increasing the pressure to the balancing point. Under such conditions, no matter how the quantities of gas change in the course of the reaction, pressure remains the same and only volume changes. The reaction takes place at constant pressure and changing volume.

Now a mole of any gas takes up a fixed volume (which happens to come to 22.4 liters) at normal atmospheric pressure, for there are always so many molecules of gas per liter at a given pressure regardless of how large or small the molecule is. In the reaction of hydrogen and oxygen



we have two moles of hydrogen reacting with one mole of oxygen to form two moles of water (and, of course, at the temperature of the reaction, the water is in the form of steam—a gas). The original gas mixture contains three moles all together, the water that is formed only two moles. The volume therefore decreases by one third. (The mass doesn't change; it's just that the atoms are three to a molecule and therefore more compactly arranged in water than in the original mixture, where they are only arranged two to a molecule.)

In a piston chamber this volume decrease would actually be reflected by a sinking of the piston and pressure would remain constant. In the bomb of a calorimeter, where the volume cannot decrease, it would be the pressure instead that would decrease by one third. (Perhaps this will be clearer to you if you study Fig. 9.)

The change in energy actually measured is not the same in these two cases. It is larger when measured at constant pressure than when measured at constant volume. The symbol for energy change at constant pressure is ΔH , where the H stands for "heat content," but is sometimes called "enthalpy." Thus, for a given reaction ΔH is greater than ΔE . The relationship can be expressed mathematically and it is possible to calculate one from the other, without too much trouble.

In nature, reactions usually take place under constant pressure; that of the natural air pressure, which is virtually constant. For that reason, ΔH is the more useful of the two

values to the chemist, and when ΔE is obtained by calorimetry it is converted to ΔH .

The question also arises as to the quantity of material for which to calculate ΔH . Obviously, if ten grams of hydrogen are burned in oxygen, ten times as much heat will be developed as would be if one gram of hydrogen were burned. Furthermore, how ought one to compare the heat developed by burning hydrogen with that developed by burning carbon? If one compared the burning of one gram of hydrogen with that of one gram of carbon, there would be the important point that one gram of hydrogen contained twelve hydrogen atoms for every carbon atom in the one gram of carbon.

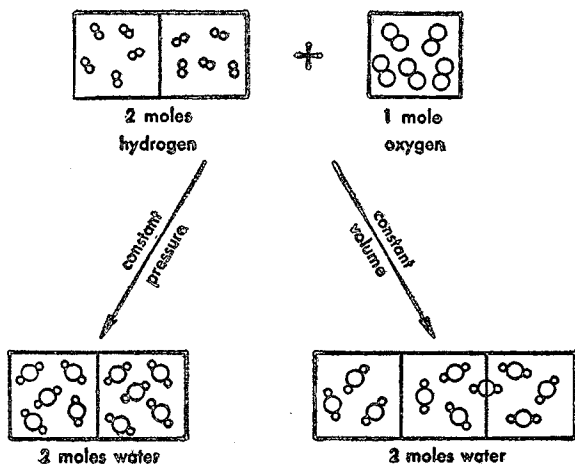
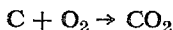


Figure 9. *Hydrogen and Oxygen Combination under Constant Pressure and under Constant Volume*

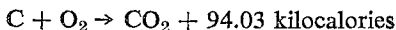
Chemists have decided that the most useful results are obtained when they determine the heat of reaction on a molar basis; that is, when they compare the heat of reaction of one mole of hydrogen burning in oxygen with one mole of carbon burning in oxygen. The weights may then be different but the same number of molecules are involved and that is what counts.

Therefore, ΔH generally stands for the "molar heat of reaction at constant pressure." This is the change in heat content involved when one mole of the substance in which we happen to be interested takes part in a reaction.

To get down to cases, consider once more the combination of carbon and oxygen:



The molar heat of reaction involved is 94.03 kilocalories. In order to make the equation complete, then, this should be included, for the combination of carbon and oxygen produces not only carbon dioxide but also energy in the form of heat and light. Therefore the equation could be written:



This can be read: one mole of carbon plus one mole of oxygen yields one mole of carbon dioxide plus 94.03 kilocalories, or (which is the same thing) 12 grams of carbon plus 32 grams of oxygen yield 44 grams of carbon dioxide plus 94.03 kilocalories.

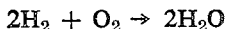
More commonly, the heat of reaction is not included in the equation itself, but is put to one side as a separate entity, thus:



Notice that ΔH is given as a negative quantity. There is some argument about this. When carbon and oxygen combine, the chemical energy of the system decreases, but the heat content increases. If you concentrate on chemical energy, you think of the energy change as negative, while, if you concentrate on heat, you think of it as positive. The number is the same in either case and it doesn't matter whether you use positive or negative, provided you stick to the same one all the time and don't switch. For a while, some chemists used the positive sign for a reaction that gave off heat, as is the case in combustion, while others used the negative sign. This led to confusion and right now the general convention is to use the negative sign in this case.

Next, consider the burning of hydrogen. This can be ex-

pressed in two ways, both involving balanced equations, as I explained in the previous chapter:



Chemists are usually interested in the heat developed by the burning of one mole of hydrogen, so the second equation, which involves one mole of hydrogen, is the one they use. (For almost any chemical purpose, other than calculating the heat of reaction, it is the first equation that is preferable.)

Very well, then; from calorimetric measurements, we can write:



(This is the heat of reaction where the water formed is allowed to cool to the liquid state, giving off the latent heat of vaporization. If the water is maintained as steam, the latent heat of vaporization cannot be added to the total and ΔH is only equal to -57.8 kilocalories. However, refinements such as this, though of much concern to chemists, need not trouble us.)

When we compare the heat of reaction for the burning of carbon with that for the burning of hydrogen, we are comparing it mole for mole. One mole of burning carbon develops nearly half again as much heat as one mole of burning hydrogen. However, a mole of carbon weighs 12 grams while a mole of hydrogen weighs but 2 grams. If we were to compare the two substances by the gram, we would find that a gram of carbon, in burning, would yield $94.03/12$ or about 7.8 kilocalories, while a gram of hydrogen, in burning, would yield $68.37/2$ or 34.2 kilocalories. On a weight-for-weight basis, then (which is the natural basis for comparison as far as the average non-chemist is concerned), burning hydrogen would yield almost $4\frac{1}{2}$ times as much energy as burning carbon. This is an example of how important it is, in making comparisons, to know exactly what it is that you are comparing.

Coal is almost pure carbon, and hydrogen is, of course, hydrogen. There are important fuels, however, with molecules containing more than one kind of element. The various pe-

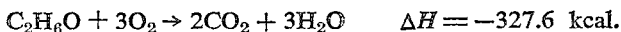
troleum fractions—gasoline, kerosene, fuel oil—have molecules containing both carbon and hydrogen. Wood is made up of molecules containing carbon, hydrogen—and oxygen as well. Such molecules, though more complex, introduce no new principles. As an example consider methane, a gas that makes up the major portion of the “natural gas” used by many in gas ranges and gas furnaces to cook food and heat houses. The methane molecule is made up of an atom of carbon and four of hydrogen (CH_4). The equation representing its combination with oxygen is:



As you see, both carbon dioxide and water are formed (and if you check the equation, you will see that it is balanced).

On a molar basis, methane yields more heat than either carbon or hydrogen. The molecular weight of methane is, however, 16 (the sum of 12 for the carbon atom and 4 for the four hydrogen atoms). One gram of methane, then, on burning yields $210.8/16$ or 13.2 kilocalories. This is more than the heat liberated by the burning of a gram of carbon and less than that liberated by the burning of a gram of hydrogen. This is to be expected, after all, of a material that is part carbon and part hydrogen.

An example of a still more complicated molecule is that of ethyl alcohol (the familiar substance contained in the beer, wine, and whiskey some of us drink on occasion), which consists of two carbon atoms, six hydrogen atoms, and an oxygen atom ($\text{C}_2\text{H}_6\text{O}$). The equation for its combustion is:

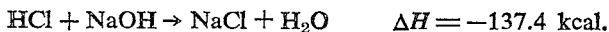


The molecular weight of ethyl alcohol is 46 (obtained by adding 24 for the two carbons, 6 for the six hydrogens and 16 for the oxygen), so that one gram of ethyl alcohol on burning yields $327.6/46$ or 7.1 kilocalories. This is far less than the heat liberated by the burning of a gram of hydrogen or of methane. It is even less than the heat liberated by the burning of a gram of carbon. The reason for this is that an oxygen atom already exists in the alcohol molecule, so that the molecule may be looked upon as already partly

burned, so to speak. The carbon and hydrogen atoms in the molecule are already combined with one of the seven oxygen atoms required all together for the formation of carbon dioxide and water.

Chemical changes need not involve oxygen in order to possess a heat of reaction (although those involving oxygen were the first to obtrude themselves on the notice of mankind). In fact, any chemical change has some heat of reaction. In the previous chapter, for instance, I gave equations representing the combination of hydrogen and chlorine to form hydrogen chloride, and the combination of sodium and chlorine to form sodium chloride. The ΔH associated with the former is -22.06 kilocalories (for the formation of one mole of hydrogen chloride out of half a mole of hydrogen and half a mole of chlorine); while that associated with the latter is -98.3 kilocalories (for the formation of a mole of sodium chloride out of a mole of sodium and half a mole of chlorine).

Again, a type of compound called an "acid" (with a certain characteristic set of properties) will react with another type of compound called a "base" (with a characteristic set of properties generally opposite to those of acids) to form compounds with neither set of properties. Such a reaction is called a "neutralization" and is exemplified by the reaction of hydrogen chloride (HCl), which, in solution, is an acid, with sodium hydroxide (NaOH), which, in solution, is a base, to form sodium chloride and water. The reaction is:



Even a change that is not usually thought of as chemical, such as the solution of a substance in water, involves a heat of reaction. This is logical, for once the compound is dissolved, the forces that held molecules of the compounds together in solid form are replaced by forces holding the molecules to nearby water molecules. This must involve some change in chemical energy. Thus, if sodium carbonate ("washing soda") is dissolved in water, the ΔH of the process (which, in this case, is called the "molar heat of solution" rather than the "molar heat of reaction") is about -5.64 kilocalories.

But this is for the solution of a mole of sodium carbonate. The formula of the substance is Na_2CO_3 and its molecular

weight is 106 (46 for the two sodium atoms, 12 for the carbon and 48 for the three oxygens), so that the solution of a gram of sodium carbonate would evolve $5.64/106$ or 0.053 kilocalories. This is 53 calories, or less than the heat evolved by the freezing of a gram of water, so it is no wonder that we are not particularly aware of heat being formed when we dissolve sodium carbonate.

Changes in chemical energy must fit the laws of thermodynamics. I implied as much when I assumed that a change in chemical energy must be balanced by an opposite change in heat energy, as otherwise the first law would be violated. However, there is more to it than that.

The earliest generalization about chemical energy was made by Lavoisier, along with another French scientist, Pierre Simon de Laplace, who in 1780, on the basis of comparatively crude (by later standards) measurements, decided that the heat absorbed in the decomposition of a compound into its elements was equal to the heat evolved by the formation of a compound from its elements.

An example will show what this means. I have already stated that the combination of hydrogen and oxygen yields 68.37 kilocalories of heat per mole of hydrogen burned or (which is the same thing if you will look at the equation) per mole of water formed:



Therefore, by the Lavoisier-Laplace generalization, the decomposition of one mole of water into hydrogen and oxygen (which can be represented by merely reversing the above equation) requires an input of 68.37 kilocalories of energy, thus:



Now the ΔH is positive because the water gains chemical energy in the course of the reaction at the expense of heat energy or (much more often) electrical energy which must be added to the water to make the reaction proceed.

We can make an analogy between chemical energy and mechanical energy. In the case of mechanical energy, the kinetic energy of an object moving upward is converted

entirely to potential energy at the height of its rise and is then reconverted to the original kinetic energy at the bottom of its subsequent fall. In the same way, the non-chemical energy added to water is converted into chemical energy by its decomposition to the hydrogen/oxygen mixture and that is reconverted to non-chemical energy when the hydrogen and oxygen combine again to form water (see Fig. 10).

(The balance, of course, is perfect only under ideal situations. In the case of mechanical energy, the absence of friction and air resistance must be assumed. In the case of chemical energy, the absence of heat loss must be assumed. In actual fact, when water is decomposed by an electric current, some of the electrical energy is always lost as heat; and if heat energy is used to decompose water, some is always lost to the outside environment. In either case, more energy is put in at the beginning than comes out at the end, but this only represents the gain of entropy and shows that, where chemical energy is concerned, the second law of thermodynamics holds as well as the first.)

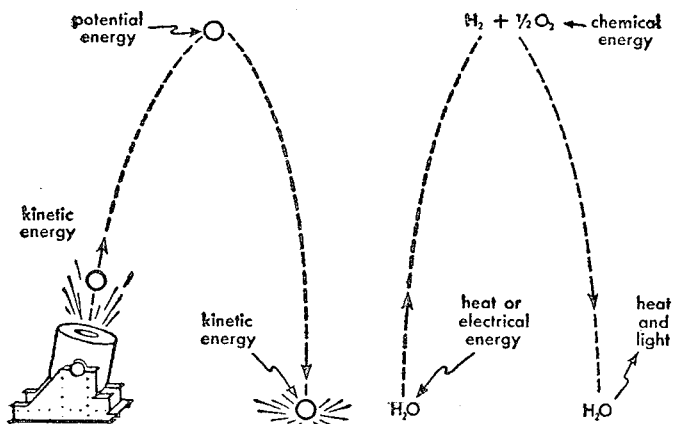


Figure 10. *Mechanical Energy and Chemical Energy*

Another important generalization was contributed in 1840 by a physicist named Germain Henri Hess, working at St. Petersburg in Russia. On the basis of his studies of heats of reaction he decided that in general, if a reaction is carried out in stages, the sum of the heats of reaction involved

in each stage is equal to the heat of reaction that is obtained if the reaction is carried out completely in one step. This is called "Hess's law" and is so important that Hess might be considered the father of "thermochemistry" ("heat chemistry").

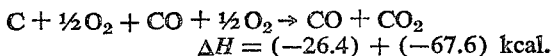
As a simple example of what I mean (but *not* one of the examples actually observed by Hess in formulating his generalization), let's consider the combustion of carbon again. It is possible for carbon to combine with a single atom of oxygen to form "carbon monoxide" (CO). We can therefore write the following equation:



Carbon monoxide is capable of burning; that is, of combining further with additional oxygen to form the more usual carbon dioxide:



If we add these two equations (just as though they were algebraic equations) we must also, by Hess's law, add the heats of reaction.



If we continue to work with techniques analogous to those used in algebra, we can cross out any symbol that appears on both sides of the equation, in this case the CO. Adding the two half moles of oxygen to form one mole, we end with:



which is exactly the heat of reaction we would observe if we burned the carbon straight through to the carbon dioxide stage in the first place.

Another way (and a more general one) of expressing Hess's law is this: If we pass from substance *A* to substance *B* by way of chemical changes, it doesn't matter what the details of the chemical changes are, the resulting over-all energy change is the same. (This is analogous to mechanical

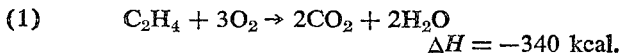
energy again, where a body in passing from one position to another changes its potential energy by a fixed amount regardless of the route it took in bringing about that change in position.)

Now although I have just used the combustion of carbon to explain Hess's law in as simple a fashion as I could, I have, in actual fact, cheated. One cannot measure, directly, the heat evolved by the burning of carbon to carbon monoxide because it is impossible to combine carbon and oxygen in such a way that only carbon monoxide is formed. Carbon monoxide *can* be formed, but not in that way.

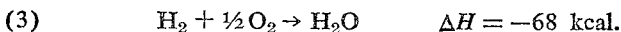
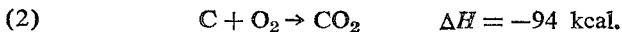
However, the thermochemist can convert carbon to carbon dioxide and measure that heat of reaction as -94.0 kilocalories. He can also burn carbon monoxide to carbon dioxide and measure the heat of reaction as -67.6 kilocalories. Then, relying on Hess's law, he can subtract the second heat of reaction from the first and deduce that the molar heat of reaction involved in the conversion of carbon to carbon monoxide is -26.4 kilocalories, even though he cannot conduct that reaction and measure the ΔH directly.

Hess's law is immensely important simply because it means that all sorts of heats of reaction that can never be directly observed can be deduced from other heats of reaction that can be observed.

For instance, there is a hydrocarbon called ethylene, which is a gas made up of molecules containing two atoms of carbon and four of hydrogen (C_2H_4). A mole of it will burn in oxygen to yield 340 kilocalories. The equation (and I shall number it and those following to avoid confusion) is:



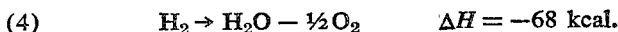
I have already discussed the burning of carbon and hydrogen, but I shall write down the equations again because we shall be using them:



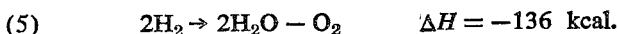
Now let us treat these chemical equations by the techniques used in algebra. (If you are one of those who are

made physically ill by algebra, I suppose you may skip the rest of the chapter without harm; otherwise, read on, and you may be amused.)

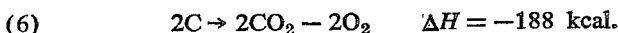
In equation (3) we can shift the $\frac{1}{2}\text{O}_2$ to the other side of the equation, changing the plus sign to a minus sign in doing so (this does not change the sign of the heat of reaction):



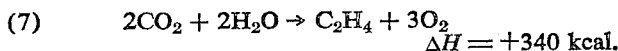
Next we double each quantity in equation (4), which also doubles the value of the heat of reaction:



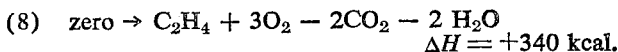
We repeat this process for equation (2), shifting the O_2 to the other side (with a change in its sign) and doubling all the terms including the heat of reaction:



In the case of equation (1), we reverse the equation, without changing any signs, and this procedure does change the sign of the heat of reaction, according to the Lavoisier-Laplace generalization:

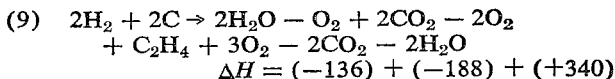


Now in equation (7) we shall shift the 2CO_2 and the $2\text{H}_2\text{O}$ to the other side of the equation, *with* a change in sign so that the heat of reaction remains positive:

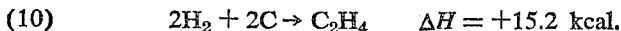


(I am writing "zero" instead of 0, because I don't want the zero confused with the symbol for oxygen.)

The next step is to add equations (5), (6), and (8), remembering also to add the heats of reaction. The result is:



If you inspect the right-hand side of equation (9), you will find that everything cancels except for the C_2H_4 so that the equation can be written more simply:



We have thus found out that if two moles of hydrogen could be made to combine with two moles of carbon to form one mole of ethylene, 15.2 kilocalories of energy would have to be added to the system to make the reaction go. Or, to put it in another way, a mole of ethylene contains 15.2 kilocalories more chemical energy than do the carbon and hydrogen that make it up. Or, to put it still another way, the "molar heat of formation" of ethylene is +15.2 kilocalories.

Actually, we cannot measure this heat of formation directly, for we cannot combine carbon and hydrogen in such a way as to form ethylene in a single step. Nevertheless, chemists feel certain of this value. In the first place, all the algebraic juggling is justified by the laws of thermodynamics, which all chemists accept. In the second, whenever such algebraic juggling can be checked by direct experiment, it has been shown to be valid.

For this reason the heats of formation of ethylene and many other compounds as well are calculated on paper to the complete satisfaction of chemists, and the fact that they cannot be directly checked by experiment and may never be, either, does not bother anyone.

Chapter 9

THE DIRECTION OF REACTION

Hess's law is an expression of the first law of thermodynamics as applied to chemical reactions, and now we must go further. Since, some chapters ago, I began to try to define life in terms of entropy, and since we suspect that

life is similar in at least some ways to combustion and therefore must involve chemical energy—we must try to find out just how the second law applies to chemical reactions.

We already know that, in the case of mechanical energy, the second law requires that objects move spontaneously from a position of high potential energy (“gravitational potential,” we might call it) to one of low, and not vice versa, so that water will flow downhill of its own accord, but not uphill. In the case of heat energy, the second law requires that heat flow spontaneously from a position of high temperature (or “heat potential”) to one of low and not vice versa, so that heat will flow from hot bodies to cold and not the other way. By analogy, then, it would seem that chemical reactions ought to take place in such a way that substances will move from a position of high “chemical potential” to low, and not vice versa.

The question is, then: What might chemical potential be and how might it be measured?

Berthelot, whom I mentioned in the previous chapter as having made many measurements of heats of reaction, thought that the answer lay in those very measurements. He felt that those chemical reactions were spontaneous which resulted in a decrease of chemical energy and the consequent appearance of heat energy; those in which, in other words, ΔH is negative. On the other hand, a reaction in which chemical energy increases so that heat is withdrawn from the environment (where ΔH is positive, in other words) will *not* progress spontaneously.

Berthelot called reactions in which ΔH is negative, and which therefore liberate heat to the environment, “exothermic” (from Greek words meaning “heat outward”). Reactions in which ΔH is positive and which absorb heat from the environment are “endothermic” (“heat inward”). What Berthelot was maintaining, then, was that only exothermic reactions were spontaneous.

There are many examples that would seem to indicate that this might be a valid generalization. For instance, a mixture of hydrogen and oxygen will, when ignited, combine rapidly and even explosively to form water, and this is an exothermic reaction. On the other hand, water can sit around forever without exhibiting any noticeable tendency to split up into hydrogen and oxygen—a process that would be endothermic if it took place. To be sure, if the water were

heated sufficiently, it would begin to split up, but that is like saying that water can flow uphill if it is pumped.

Of course, you might point out that a mixture of hydrogen and oxygen will also sit around forever without showing any tendency to react unless it, too, is subjected to heat. However, there is this difference. Once a mixture of hydrogen and oxygen is ignited by the small addition of heat, the reaction will carry on without further interference for an indefinite time thereafter. If, on the other hand, the addition of intense heat, or of an electric current, begins to bring about the decomposition of water to hydrogen and oxygen, the withdrawal of that heat or that current, after no matter how long a period of application, would bring the decomposition to an instant halt.

To summarize, then, the exothermic union of hydrogen and oxygen can be made to proceed spontaneously, but the endothermic splitting of water cannot be.

Unfortunately for Berthelot, however, we have here a case of a generalization that is simply not general enough. All sorts of exothermal reactions do indeed take place spontaneously and all sorts of endothermal reactions do indeed not take place spontaneously. However, there are some reactions that are endothermal and yet do take place spontaneously. The best examples include the solution of certain substances in water. If ammonium nitrate is added to water it dissolves spontaneously, yet the ΔH involved in the process is positive. Chemical energy is gained and heat energy is withdrawn from the water, which grows icy to the touch. No matter how many cases there are that uphold Berthelot's generalization, that one exception is enough to invalidate it.

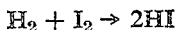
But, as a matter of fact, the insufficiency of Berthelot's generalization is not highlighted merely by a few exceptions. There are more subtle objections involved and this brings us to the concept of "chemical equilibrium."

In the reactions that are of most interest to thermochemists, those involving combustion, the direction of reaction seems "one-way." That is, carbon and oxygen will combine completely to form carbon dioxide; hydrogen and oxygen will combine completely to form water; compounds with molecules containing both carbon and hydrogen will combine completely with oxygen to form both carbon dioxide and water. In no case, moreover, will water measurably

split to hydrogen and oxygen, or carbon dioxide split to carbon and oxygen; nor will water combine with carbon dioxide to form ethylene or any other hydrocarbon. Such "one-way" reactions are said to be "irreversible."

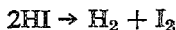
This is not the situation, however, with respect to many other chemical reactions. An example of what I mean involves hydrogen and iodine. Iodine is a slate-gray solid at ordinary temperature, but gentle heating will produce a violet vapor and at temperatures above 184° C. it exists only as a vapor. In the gaseous state, iodine consists of molecules made up of two atoms apiece (I_2).

If, then, at a temperature of, say, 445° C., hydrogen and iodine are mixed, there is a combination of the two and hydrogen iodide, HI, is formed:



However, not all the hydrogen and iodine combine. No matter how long a time is allowed to elapse, some hydrogen and some iodine remain uncombined.

If, on the other hand, a sample of pure hydrogen iodide is obtained and this is heated to 445° C. it decomposes, forming a mixture of hydrogen and iodine:



Here again, no matter how long a time is allowed to elapse, not all the hydrogen iodide breaks up. In fact, whether you start with a mixture of hydrogen and iodine, or with hydrogen iodide, you end up with precisely the same situation. About 80 per cent of the hydrogen and iodine is in the combined form as HI; about 20 per cent is not. Such a reaction, which can go either way, is a "reversible reaction."

Apparently what happens is this. Hydrogen and iodine, mixed at a high temperature, combine rapidly to form hydrogen iodide. The hydrogen iodide that is formed tends to break down at that temperature, but so much hydrogen and iodine are combining and there is so little hydrogen iodide, just at first, to break down, that the combining effect is preponderant and hydrogen iodide accumulates.

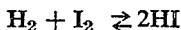
As hydrogen iodide accumulates, however, more and more of the accumulating molecules will break down. At

the same time, the mixture of still uncombined hydrogen and iodine is depleted, and the rate of hydrogen iodide formation declines. As the rate of hydrogen iodide formation declines and the rate of hydrogen iodide breakdown increases, there must come a point where the two are in balance. That point is reached under the conditions of the experiment, when 80 per cent of the hydrogen and iodine have combined. Nothing further seems to happen, *not* because nothing is happening in actuality, but because two opposing effects are taking place at the same rate.

Now suppose you begin with pure hydrogen iodide (see Fig. 11). Some begins to break up at once. As hydrogen and iodine accumulate through the breakdown there is a tendency to recombine again, which increases in rate as the two gases accumulate. Conversely, as the number of hydrogen iodide molecules decreases through breakdown, the rate at which the remainder breaks down must decrease. Again, there comes a point of balance, and the same point of balance, experimental observation shows, is reached.

When two opposing reactions are equal in rate so that there is no longer any over-all change, the situation is said to be one of "chemical equilibrium." The word "equilibrium" (from Latin words meaning "equal balance") brings up the picture of a balance holding level and motionless under balanced forces. This is actually "static equilibrium." Chemical equilibrium is, however, not static. Two opposed reactions are both moving rapidly, but happen to negate each other. This is "dynamic equilibrium."

To indicate a situation of chemical equilibrium, an equation for the reaction of hydrogen and iodine is usually written as follows, with two oppositely pointed arrows to indicate that there is movement in each direction:



Any number of other reactions may be written in the same way.

The crucial point is this: Any reaction has, under fixed conditions, a particular ΔH . The reverse reaction, under those same conditions, has the same ΔH with the opposite sign, by the Lavoisier-Laplace generalization and, which is more important, by the requirements of the first law. There-

fore, in *any* reversible reaction, if one reaction is exothermic, the reverse reaction must be endothermic. In the case

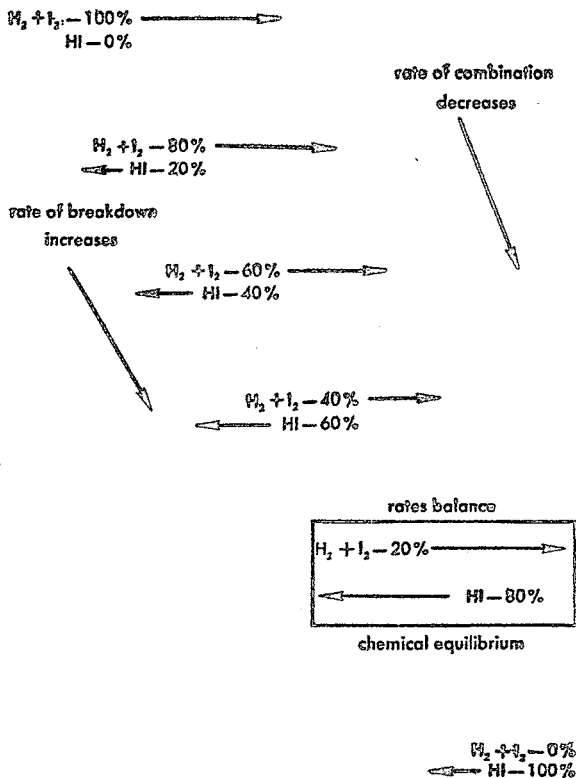


Figure 11. *The Relationship between the Rate of Reaction and Chemical Equilibrium*

of hydrogen, iodine and hydrogen iodide, the breakup of hydrogen iodide to hydrogen and iodine is exothermic, with a ΔH of -6.0 kilocalories. Therefore, the combination of hydrogen and iodine is endothermic, with a ΔH of $+6.0$ kilocalories.

If Berthelot's suggestion, that only exothermic reactions were spontaneous, were valid, then, left to itself, hydrogen

iodide would break down and would not re-form. This, however, is at variance with actual observation. Hydrogen iodide *does* re-form; in the case cited it actually forms to a greater extent than it breaks down.

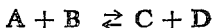
If Berthelot were correct, there would be no such thing as a reversible reaction; all reactions would go in one direction only. Since this is certainly not in accord with observation, Berthelot's suggestion was objected to and forced into abandonment almost as quickly as it was made. (In fact, I wonder Berthelot did not see this himself and avoid making the generalization in the first place—but then, it is easy to be wise after the fact.)

This leaves us still in search of our chemical potential. Well, one factor that does seem to have some influence in the direction taken by a reversible reaction is the quantity of each material present. If there is a sufficient preponderance of hydrogen and iodine, the over-all change is in the direction of combination. If there is a sufficient preponderance of hydrogen iodide, the over-all change is in the direction of breakdown. If equilibrium has already been reached, it can be disrupted and the reaction sent hurtling in either direction by the simple expedient of adding more hydrogen iodide or of adding more of a mixture of hydrogen and iodine.

The first to note this fact clearly was the French chemist Claude Louis Berthollet in 1803. (This man is not to be confused with the Berthelot discussed in the previous section, who lived a half century later.)

For over half a century Berthollet's hypothesis concerning the influence of mass on the direction of reaction did not win acceptance though chemist after chemist fumbled with the notion. Finally, in 1863, the Norwegian chemists Cato Maximilian Guldberg and Peter Waage worked the matter out thoroughly and announced what is now called the "law of mass action." (They published in Norwegian, unfortunately, and there was a delay of about fifteen years before the French and German leaders in the field discovered the paper.)

To show the significance of this law most clearly, let's consider a generalized reversible chemical reaction, one in which there is no need to specify the actual substances involved. Such a reaction can be written:



If we consider the reaction as going to the right, then the substances A and B are combining to form C and D. However, in order to do so, they must find each other, so to speak. The more there is of these two substances, the easier it is for them to find each other and the faster the combination proceeds. Nor is it so much the total quantity that counts as the quantity present in a given volume; quantity per volume being called the "concentration." (As an analogy, a young lady with matrimonial designs may find more men, all told, in rural Nevada than in the city of Holyoke, Massachusetts, but in Nevada they are spread out thinly. The concentration is greater in Holyoke, and the chances of marriage are likewise greater.)

If the concentration of A is doubled, the number of meetings of A and B and, consequently, the rate of combination is doubled. The same is true if the concentration of B is doubled. If the concentrations of both A and B are doubled, then the meetings and the rate of combination are doubled twice, or increased fourfold. The rate of combination is thus proportional to the concentration of A multiplied by the concentration of B. It is customary to express the concentration of any substance by enclosing the symbol for that substance in square brackets, so we can put what I have just said into symbols as follows:

$$\text{reaction rate to the right} \sim [A] [B]$$

Now whenever any quantity, x , is directly proportional to any other quantity, y , the proportionality can be changed to an equality if y is multiplied by some fixed value usually represented by k and called the "proportionality constant." In other words if $x \sim y$, then $x = ky$.

Applying this to the case of the reaction to the right:

$$\text{reaction rate to the right} = k[A] [B]$$

The actual value of k in some particular reaction must be determined by actual experiment. Right now, though, we don't need the actual value. It is enough to know that k is some fixed value.

If we next concentrate on the reaction to the left, which

is the reverse of the reaction to the right, we see that this involves the combination of C and D to form A and B. By the same line of reasoning as before, this leftward reaction rate depends on the concentrations of C and D, and we can say:

$$\text{reaction rate to the left} = k'[\text{C}][\text{D}]$$

where k' represents another proportionality constant, presumably different in value from the first and therefore distinguished by the prime (') superscript.

Now at the equilibrium, the reactions are equal in rate in both directions (this condition of dynamic equilibrium having been first made clear by the English chemist Alexander William Williamson in 1850) so that, *at equilibrium*:

$$k[\text{A}][\text{B}] = k'[\text{C}][\text{D}]$$

and, by ordinary algebraic manipulation:

$$\frac{[\text{C}][\text{D}]}{[\text{A}][\text{B}]} = k/k'$$

Since both k and k' are constants, the division of one by the other must yield still another constant, which we can call K . Therefore:

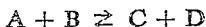
$$\frac{[\text{C}][\text{D}]}{[\text{A}][\text{B}]} = K$$

and K is called the "equilibrium constant." Its value fixes the conditions of concentration that must prevail at equilibrium.

Suppose, for instance, that at equilibrium more of C is added to the mixture. In that case the product $[\text{C}][\text{D}]$, which is the numerator of the fraction, increases. In order that K retain its fixed value, the denominator $[\text{A}][\text{B}]$ must also increase. The only way in which this can happen is for some of the C and D to combine to form A and B. This means that $[\text{C}][\text{D}]$ will decline in value a bit, while that of $[\text{A}][\text{B}]$ rises until the fraction once again has the value of K , at which point equilibrium is again reached. The same would happen if D rather than C were added.

You can check for yourself what would happen if either A or B (or both) were added, or if either or both were taken out of the reaction mixture, or if C or D or both were taken out.

To summarize, and to give you the answer, the effect of the law of mass action is to state that in the reversible reaction:



the reaction is forced to the right if A or B is added or if C or D is subtracted. It is forced to the left if C or D is added or if A or B is subtracted.

(In 1901 the American chemist Gilbert Newton Lewis modified the law of mass action by showing that it was not the concentration of substances that directed the course of a reversible reaction but a quantity he called the "activity" of those substances. This activity was related to the concentration and was often very nearly equal to it, particularly if the concentration was not high. This refinement need not concern us further, though it is well to know that it exists.)

This response of a reversible reaction to changes in concentration was made part of a still broader generalization in 1888 when the French chemist Henry Louis Le Châtelier stated: In any system at equilibrium, any change in conditions results in a shift of equilibrium in the direction that will partially nullify the change.

This generalization is referred to as "Le Châtelier's principle."

In other words, if A or B is added to the reversible reaction we have just been considering, the equilibrium shifts in the direction that uses up some of the added A or B. More broadly, if an equilibrium mixture is heated, the equilibrium shifts in such a direction that some of the heat is absorbed; that is, it shifts in the endothermic direction. Similarly, if an equilibrium mixture is cooled, the shift is in the direction of the exothermic reaction. (We can predict from this alone that heating the hydrogen, iodine, hydrogen iodide system would shift the direction of reaction toward the formation of more hydrogen iodide, while cooling it would shift it in the opposite direction.) Again, an increase in pressure would shift the equilibrium in the direction that would result in decreased volume.

In fact it is possible to view Le Châtelier's principle as of universal application, holding true even for human behavior. The average man usually maintains some sort of equilibrium between income and expenses so that neither savings nor debts accumulate to too extraordinary a degree. If his salary is raised or an uncle leaves him a legacy, his scale of living generally increases so as partially to compensate for the added income. If, on the other hand, his salary is cut, or he loses his job, he begins to economize, cutting his expenses and thus partially compensating for the decreased income.

After the law of mass action was accepted, it was still necessary to tie it to the laws of thermodynamics in order that chemical energy might be shown, quite clearly, to be following the same rules that other forms of energy do.

As early as 1869 a German chemist, August Friedrich Horstmann, applied the notions of thermodynamics to the decomposition of a compound called ammonium chloride, and through the 1870s and 1880s a number of other chemists tackled the problem.

The most important work was done by an American theoretical chemist, Josiah Willard Gibbs. In 1876 and 1878, Gibbs, working at Yale, published papers in the *Transactions of the Connecticut Academy of Sciences* in which the laws of thermodynamics were applied to chemical reactions in great detail. The law of mass action was shown by him to be an inevitable consequence of the laws of thermodynamics. The treatment was so thorough, in fact, that the field of chemical energy was no longer merely "thermochemistry" but could be considered "chemical thermodynamics."

Unfortunately for Gibbs, the current of theoretical science flowed most strongly in Europe in those days, and American scientists were little regarded. Since American scientific journals were not read closely, Gibbs's work escaped the notice of European scholars for twenty years.

Finally, in the 1890s, Le Châtelier himself translated Gibbs's papers into French, while the German physical chemist Wilhelm Ostwald translated them into German. Then, and only then, did Gibbs receive his due. By then, Helmholtz (of first-law fame) and the Dutch chemist Jacobus Hendricus van't Hoff had established the principles of chemical thermodynamics independently, but Gibbs was recognized as

having gotten there first. In 1901, Gibbs received official European recognition of his feat by being awarded the Copley medal of the Royal Society of London, and none too soon, either, for he died in 1903.

In essence, Gibbs's treatment revolved about the fact that, in any spontaneous chemical reaction, entropy must increase, as it must in any change involving energy. Since the total energy involved could not change, and the increase in entropy is a way of measuring energy that could not be converted into useful work, then the quantity of energy that *could* be converted into useful work had to decrease.

The quantity of energy available for conversion into useful work, during the course of a chemical reaction, differs according to whether the reaction takes place at constant volume or constant pressure. As I explained in the previous chapter, systems under constant pressure more closely resemble reality. The quantity of energy available for conversion to work in the course of a chemical reaction at constant pressure is called, in Gibbs's honor, the "Gibbs function." In the early 1920s, Lewis introduced the phrase "free energy" for it and that has caught on. The result of this double name, however, is a double symbol. Free energy is sometimes symbolized as F (for free energy) and sometimes as G (for Gibbs function). Since free energy is now the common name, I prefer F and shall use it.

Gibbs showed that, in any reaction at some particular temperature (T), there was a simple relationship between the change in heat content (ΔH), the change in free energy (ΔF), and the change in entropy (ΔS). This relationship is:

$$\Delta H = \Delta F + T\Delta S$$

Now in thermodynamics, temperature is always given in the absolute scale (see Chapter 4) where there are no temperatures lower than zero. The value of T is therefore always positive. From the second law of thermodynamics we know that in any spontaneous reaction entropy must increase, so the change in entropy (ΔS) must always be positive also. Since the product of two positive quantities is always positive, we end by saying that $T\Delta S$ is always positive.

On the other hand, by Gibbs's reasoning, the free energy always diminishes in any spontaneous chemical reaction, so that the change in free energy (ΔF) must always be negative.

As you see, then, the value of ΔH is the sum of a negative number and a positive number. But though thermodynamics tells us that one term of the sum must be negative and one must be positive, it does not tell us the exact size of either. If the negative ΔF is greater than the positive $T\Delta S$ (as is quite often the case), the sum of the two (ΔH) is negative and the reaction is exothermic. If the positive $T\Delta S$ happens to be the larger (as is sometimes the case), then the sum of the two is positive and the reaction is endothermic.

In either case, though, exothermic or endothermic, entropy goes up and free energy goes down, and the reaction is spontaneous.

The thing to watch, then, is not whether heat is given off or taken up; rather one must watch the free energy. A reaction will always head in the direction of lower free energy, as heat will flow toward the cold spot and a rolling body will move toward the valley.

Gibbs went on to show that the quantity of free energy present in any chemical substance varied somewhat with the concentration of the substance. By shifting the concentrations of the substances it was thus possible to shift the nature of the free energy change so that it became negative when a reaction moved in one direction, whereas previously it had been negative when a reaction moved in the other direction. This explained why a reversible reaction could be shifted in this direction or that by the addition or removal of components of the system.

It was as though a ball were resting on a smooth, perfectly level plane (equilibrium) that could be tipped slightly, either one way or the other. The ball could be made to roll in either direction.

Gibbs worked out an expression that combined free energy and concentration and called this the "chemical potential." Here, at last, is what we were looking for in this chapter. Every reaction takes place in such a way as to decrease its chemical potential, as defined by Gibbs.

Chapter 10

THE ELECTRICAL ROUTE

In considering the effects of free energy changes on chemical systems, it simplifies matters to make the progress of a chemical reaction analogous to motion in a gravitational field. This is a reasonable thing to do since the whole point of the previous chapter was that chemical reactions move in response to chemical potential as falling bodies move in response to gravitational potential. Besides, we are so accustomed to the motion of falling bodies that it is almost second nature to visualize everything we possibly can in that form.

For instance, consider the combination of hydrogen and oxygen to form water. The mixture of hydrogen and oxygen has a certain free energy content per mole ("molar free energy") and so has the water. We can safely say, just from our knowledge that hydrogen and oxygen will combine vigorously at the least encouragement, that the free energy content of the mixture of hydrogen and oxygen is higher than that of the water that it forms on combination. We can present this in schematic form (without any attempt to use actual figures) in the accompanying diagram (see Fig. 12). Point A represents the free energy of a pure mixture of hydrogen and oxygen, and point B (considerably lower on the scale) represents the free energy of the pure water formed by a complete combination of all the hydrogen and oxygen.

As the hydrogen and oxygen combine to form water, more and more of the hydrogen and oxygen are replaced by water in this closed system we are imagining. Little by little a high-free-energy mixture is being replaced by a low-free-energy compound, so that the total free energy declines steadily.

The simplest way of showing this is to draw a straight

line from A to B in the diagram. The situation is thus made to represent an inclined plane and our experience with spontaneous motion makes the direction of reaction seem inevitable. The combination of hydrogen and oxygen to form water is a process that "rolls downhill" from a free energy standpoint. The reverse reaction, which is the decomposition of water, requires an effortful movement "uphill."

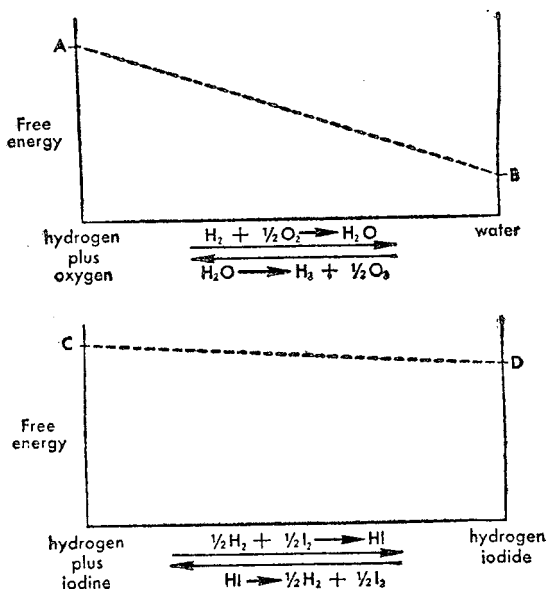


Figure 12. Free Energy Change in the Formation of Water (top) and Hydrogen Iodide (bottom)

To be sure, the straight line I have drawn would be valid only if the molar free energy of hydrogen and oxygen and water did not change with concentration. But values of the molar free energies do, however, change a bit with concentration so that the straight line is not really straight. In this particular case the effect is small enough to be ignored, but where the free energy change of a reaction is considerably smaller than in the case of the combination of hydrogen and oxygen, the effect becomes important indeed.

Consider the combination of hydrogen and iodine to form hydrogen iodide. Here the free energy change involved in the reaction is quite small. We can show this in Fig. 12 by placing the free energy mark (C) for the pure mixture of hydrogen and iodine very little higher than the mark (D) for the free energy of the pure hydrogen iodide formed by a complete combination of the hydrogen and iodine.

We might suppose, then, that we could still draw a straight line from one mark to the other, making another inclined plane; a gently sloping one, to be sure, but an inclined plane just the same. The reaction should still "roll downhill" from the mixture of hydrogen and iodine to hydrogen iodide; while the hydrogen iodide at the bottom of the hill would be unable to dissociate unless it were somehow forced "uphill." The combination of hydrogen and iodine might be less violent and more sluggish than the combination of hydrogen and oxygen because the hillside is more gently sloped but otherwise the principle, it would seem, is similar.

Yet we know that this is not what happens in actual fact. The hydrogen and iodine do not combine entirely and do not, therefore, "roll all the way downhill." Furthermore, pure hydrogen iodide will decompose to a certain extent and would therefore seem to "roll uphill." The straight-line diagram can't be right, then.

The answer is that molar free energy decreases with concentration. If the concentration of hydrogen plus iodine is cut by 10 per cent through the formation of hydrogen iodide, the free energy of the mixture is not cut by a mere 10 per cent but by something more than 10 per cent. This means that the free energy line, starting at point C, dips below the path followed by the straight line.

The same is true in the case of hydrogen iodide. If some of it breaks down, what is left has a smaller free energy than you would expect. The free energy line, starting at point D, also dips below the straight line.

In fact, the line representing the free energy would be hammock-shaped, as shown in Fig. 13, and most of it would represent a total free energy content that is less than that at *either* extreme.

This should be stressed. Because of the fact that molar free energy varies with concentration, the free energy of a given reaction does not reach its minimum at one of the extremes, with the reaction at completion in one direction

or the other. Instead it reaches a minimum at some intermediate point where the lowered concentration of all the components gives the total free energy an extra push downward, so to speak.

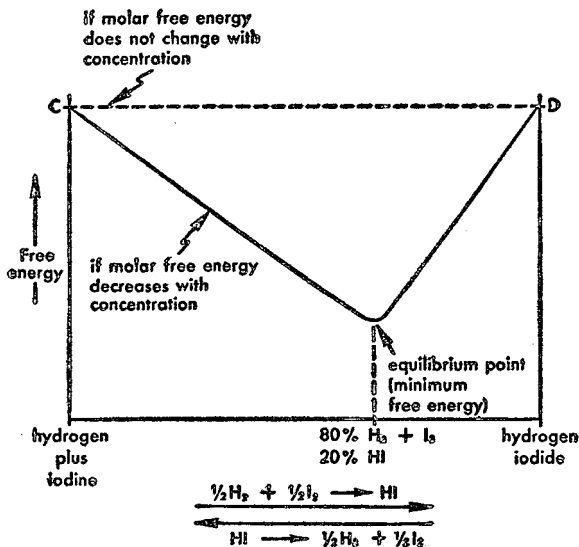


Figure 13. Free Energy Change and Equilibrium

In consequence, a reaction, coming to a halt at the point of minimum free energy content, finds itself at some intermediate point, which becomes the point of equilibrium. What's more, the reaction comes to a halt at that same intermediate point of minimum free energy whether you start at one extreme (with a pure mixture of hydrogen and iodine) or at the other (with pure hydrogen iodide).

The analogy with motion under gravity involves, now, not an inclined plane but two inclined planes; or, to be even more metaphoric, two hillsides with a valley between. If a ball were released at C it would roll down the hill to the low point of the valley, while if it were released at D it would roll down the other hill to the same point.

The smaller the change in free energy between the extremes of the reacting system, the closer the point of equilib-

rium comes to the midpoint of the reaction. Where the free energy difference between the extremes is high, the point of equilibrium shifts far over toward the extreme with lower free energy.

In the case of the combination of hydrogen and oxygen, the equilibrium shifts so far in the direction of water that the difference in position between the equilibrium point and that of pure water is immeasurable. This equilibrium point would have a free energy value lower than that of pure water but so microscopically lower that the difference could not be measured (though it could be calculated from free energy data).

This means that hydrogen and oxygen, in a closed system, would never entirely combine to form water, but that there would always be an immeasurably small quantity of hydrogen and oxygen left uncombined. Furthermore, if you began with pure steam, under the same conditions of temperature and pressure, an immeasurably small quantity of it would spontaneously decompose to hydrogen and oxygen.

From the standpoint of thermodynamics, then, all chemical reactions are reversible. In a practical sense, however, we can continue to view as irreversible those reactions with equilibrium points so close to one extreme as to be virtually indistinguishable from it.

Since the position of equilibrium depends on the free energy change involved in going from one side of a chemical equation to the other, it is possible to calculate that change quite simply from the equilibrium constant. The catch, however, is that the equilibrium constant is not always easy to determine; certainly not in the reactions usually considered irreversible. Other methods of calculating free energy changes have been searched for, therefore, and found. One of these, which is worth considerable discussion because it illuminates important corners of chemistry, involves electricity.

To discuss this, however, requires a new and deeper look at the atom. So far, in talking about kinetic theory of gases and about chemical reactions, I have treated atoms as though they were tiny billiard balls. Actually, however, there is far more to them than that. Atoms are composed of still smaller particles, a discovery that first broke on the world of science with revolutionary impact in the 1890s.

There are three chief varieties of these "subatomic particles": protons, neutrons, and electrons. Of these, the protons and neutrons are comparatively massive and are tightly crammed into a tiny package (the "atomic nucleus") at the very center of the atom. They are out of the way and are not involved in ordinary chemical reactions.

All about the nucleus, filling out the remainder of the atom, are the electrons, which are very light particles. The number of electrons per atom is equal to the number of protons in the nucleus of that atom, and this number is different for the atoms of each element. (It is called, in fact, the "atomic number.") The simplest atom, that of hydrogen, contains only one proton and one electron, but other atoms contain more. Some of the complex atoms created in recent years in the laboratory contain more than a hundred. For future reference, the atomic number of some of the more familiar elements is listed in Table 2.

Now each electron has a fixed "electric charge,"¹ and so has each proton. (The neutrons have no charge and, while extremely important, will because of this play no further part in the line of exposition to follow.) The quantity of charge is exactly the same on both electron and proton but the nature of the charge is different. This can be shown by the fact that a proton distinguishes between an electron and another proton; it attracts the former and repels the latter. Similarly, an electron will attract a proton and repel another electron.

It has been decided then, quite arbitrarily, to label the electric charge on the electron as negative and that on the proton as positive. Each electron has a charge of -1 , therefore, while each proton has a charge of $+1$, and the rule is set up that "like charges repel; unlike charges attract." (Giving each particle a unit charge is a good idea, too, for as far as we know, no smaller charge exists; and, very probably, no smaller charge can exist.)

In the ordinary atom, I have said, the number of protons within the central nucleus (again, the "atomic number") is

¹ Exactly what the charge *is*, is not known, but this can be left to philosophers. The charge, whatever it is, is a property that evidences itself as an attraction or repulsion that can be accurately measured and that follows set rules. This is sufficient for physicists.

Table 2
THE ATOMIC NUMBER

Element	Atomic Number	Element	Atomic Number
Hydrogen (H)	1	Cobalt (Co)	27
Helium (He)	2	Nickel (Ni)	28
Carbon (C)	6	Copper (Cu)	29
Nitrogen (N)	7	Zinc (Zn)	30
Oxygen (O)	8	Silver (Ag)	47
Fluorine (F)	9	Tin (Sn)	50
Sodium (Na)	11	Iodine (I)	53
Magnesium (Mg)	12	Tungsten (W)	74
Aluminum (Al)	13	Platinum (Pt)	78
Phosphorus (P)	15	Gold (Au)	79
Sulfur (S)	16	Mercury (Hg)	80
Chlorine (Cl)	17	Lead (Pb)	82
Potassium (K)	19	Radium (Ra)	88
Calcium (Ca)	20	Thorium (Th)	90
Chromium (Cr)	24	Uranium (U)	92
Manganese (Mn)	25	Plutonium (Pu)	94
Iron (Fe)	26	Lawrencium (Lw)	103

NOTE: All the numbers skipped in the list above are represented by known elements, so that there are a total of 103 elements now known.

exactly balanced by the number of electrons on the outskirts, so that, on the whole, the atom is uncharged or "electrically neutral." However, while the protons, well hidden in the nucleus, are not affected by atomic and molecular collisions, the light electrons, exposed on the outskirts of the atom, bear the brunt. Chemical reactions invariably involve changes of some sort in the electron distribution of the atoms participating in the reaction. It is, in fact, the existence of such changes in electron distribution that can serve to identify a chemical reaction from other types of change.

These changes may or may not affect the electrical neutrality of an atom. For instance, two atoms, on colliding, may share electrons in order to set up a particularly stable electron distribution (something I shall be mentioning again later in the book). The shared electrons can remain shared only if the atoms remain in virtual contact. To pull the atoms apart means that a stable arrangement will be broken

up and this requires a considerable input of energy. In the absence of such an energy input, the atoms sharing electrons will remain together indefinitely and this, in fact, is the basis of the chemical bonds I have been dealing with in previous chapters. The atoms in the molecules of hydrogen, oxygen, water, carbon dioxide, hydrogen iodide, and many others, are held together by virtue of the electrons they share.

Where atoms are shared in this way, the electrical neutrality of the atoms is not affected.

However, it so happens that certain electrons in the atoms of some elements are held quite weakly. For instance, the sodium atom holds the outermost of its eleven electrons with a feeble grip. In the neighborhood of a chlorine atom (which holds all its electrons very firmly) that outermost electron of the sodium atom is not shared but is transferred bodily from the sodium atom to the chlorine atom.

Once this transfer takes place, the sodium atom is left with only ten electrons, which is one electron less than the number required to neutralize the positive charge of the eleven protons in the sodium nucleus. The sodium atom is no longer electrically neutral. It has an over-all charge of $+1$, equal, you see, to the charge of the one non-balanced proton. On the other hand, the chlorine atom now has a total of eighteen electrons, or one electron more than the number required to neutralize the seventeen protons in its nucleus. It has an over-all charge equal to that extra electron, or -1 .

Such atoms, with electric charges due to a deficiency or excess of electrons, are called "ions" and, naturally, there are "positive ions" and "negative ions." Ions are symbolized by the ordinary symbol of the element with a little plus mark or minus mark to indicate the charge. Thus, there is the "sodium ion," Na^+ , and the "chloride ion," Cl^- . (Why it should be "chloride ion" and not "chlorine ion" involves a discussion of the history of chemical terminology for which room is lacking. I can only ask the reader to accept the names of these and other ions, as given.)

Some atoms can lose two or even three electrons in the course of a chemical reaction; or can even show themselves capable of varying their loss, losing two electrons under one set of conditions and three under another set. It is even possible for a group of atoms held together by ordinary

shared-electron bonds to lose or gain one or more electrons. The constituent atoms remain together but now have an electric charge and are "compound ions." All this is not a subject into which we need go in great detail; nevertheless, for future reference, a list of the more familiar ions is given in Table 3.

Table 3

IONS

Positive		Negative	
Name	Symbol	Name	Symbol
Hydrogen ion	H ⁺	Fluoride ion	F ⁻
Sodium ion	Na ⁺	Chloride ion	Cl ⁻
Potassium ion	K ⁺	Iodide ion	I ⁻
Magnesium ion	Mg ⁺⁺	Sulfide ion	S ⁼
Calcium ion	Ca ⁺⁺	Carbonate ion	CO ₃ ⁼
Zinc ion	Zn ⁺⁺	Nitrate ion	NO ₃ ⁻
Cuprous ion	Cu ⁺	Sulfate ion	SO ₄ ⁼
Cupric ion	Cu ⁺⁺	Phosphate ion	PO ₄ ⁼
Ferrous ion	Fe ⁺⁺		
Ferric ion	Fe ⁺⁺⁺		
Ammonium ion	NH ₄ ⁺		

NOTE: When an atom or atom group gains or loses more than one electron, the fact is indicated by the presence of more than one plus or minus mark. The calcium atom loses two electrons and thus produces the doubly charged ion, Ca⁺⁺, while the sulfur, gaining two electrons, becomes S⁼. Notice also that the iron atom can lose either two or three electrons, forming either a doubly charged or a triply charged ion, the two being distinguished by the name used, which in this case, by the way, derives from the Latin *ferrum* rather than from the English "iron."

The generalization that opposite charges attract holds for ions. Once sodium ions and chloride ions are formed by a shift in an electron from a sodium atom to a chlorine atom, the two ions remain together, bound by the force of the attraction between the positive charge of one and the negative charge of the other. Together, they form sodium chloride (ordinary table salt).

Substances such as sodium chloride, which are held together by such attractions, do not consist of molecules in

the same sense that carbon dioxide or water does. Within a crystal of sodium chloride, a particular sodium ion is not permanently associated with a particular chloride ion to form a discrete two-atom particle. This shows up most clearly when sodium chloride is dissolved. In a sodium chloride solution, the sodium ions and the chloride ions move about freely. To be sure, there is an over-all attraction between positive and negative charges but a particular sodium ion can move without dragging a particular chloride ion with it. (In fact, under the influence of an electric current, the two sets of ions can be made to move in opposite directions.)

There are chemical reactions that can best be interpreted in the light of such ionic independence; that can be most clearly described by concentrating on one of the ions, for instance, and ignoring the other.

An example involves zinc (chemical symbol, Zn) and copper sulfate (with a molecule made up of one copper atom—chemical symbol, Cu, from the Latin *cuprum*—one sulfur atom, and four oxygen atoms, thus, CuSO_4). If metallic zinc is added to a solution of copper sulfate, a spontaneous reaction is initiated in which the zinc goes into solution, becoming zinc sulfate (ZnSO_4) while the copper comes out of solution as the familiar reddish metal.

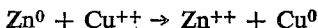
The ordinary way of writing the equation to represent this chemical reaction would be



and the picture you would get would be that of the sulfate group (SO_4) detaching itself from the copper and attaching itself to the zinc. In actual fact, however, the sulfate group does not engage in any such activity. It is the one component of the system that plays no direct role at all, in fact, and this can be shown by looking at the reaction from the ionic standpoint.

Copper sulfate is held together by ionic attraction (at least in part) and exists in solution not as discrete molecules but as a mixture of cupric ions (Cu^{++}) and sulfate ions (SO_4^{--}). What happens in the reaction, then, is that an atom of zinc (which can be written as Zn^0 , to stress the fact that in the metallic state it is an uncharged atom) loses two electrons to become a zinc ion (Zn^{++}) and that these two electrons are gained by the cupric ion, which then becomes a neutral cop-

per atom (Cu^0), in which form it settles out as the metal. The equation can be written thus:



This reaction involves a change in the nature of the ionic bond. To begin with the ionic bond was one between cupric ion and sulfate ion and in the end it was one between zinc ion and sulfate ion. Changes in ionic bonds involve changes in free energy as much as do changes in any other form of bond. In this case the reaction goes spontaneously in the direction indicated in the equation above, so we can be sure that in change from metallic zinc and cupric ion to zinc ion and metallic copper there is a decrease in free energy.

Now suppose the two reacting substances are separated. In one container you have a rod of zinc immersed in a solution of zinc sulfate. In another container you have a rod of copper immersed in a solution of copper sulfate. There is a porous partition between the two containers that will permit ions to pass through in either direction but only very slowly, under ordinary conditions. If such a system is left to itself, the reaction discussed above does not proceed at any significant rate.

What does happen, however, is that there is a certain tendency for the zinc atoms in the zinc rod to lose electrons and form zinc ions. There are no copper ions in the vicinity to take up the electrons, which therefore accumulate as "free electrons." This does not proceed very far, for the electrons that accumulate, all possessing like charge, develop a mutual repulsion so that each, in a manner of speaking, tries to get as far away as possible from all the rest, setting up a kind of "electron pressure." Additional electrons, arising from the ionization of more zinc atoms, must battle this pressure. With each electron added, the electron pressure increases until no further ionization can take place. The accumulation of electrons at that time is still extremely small. The same is happening in the copper rod, for the copper atom, too, has a tendency to lose electrons. Nevertheless, the electron pressure is greater in the zinc rod, for zinc atoms have a greater tendency to lose electrons than copper atoms do, and more electrons therefore accumulate in the zinc rod.

Suppose, now, that the zinc rod is connected to the copper rod by means of a metal wire. Here something new and interesting has a chance to happen. Ordinary atoms could not travel through the wire; the atoms already in the wire and making up its substance form an effective barrier against that. The electron, however, is much smaller than the atom and much lighter. It is only $1/1837$ as massive as even the lightest atom (which is hydrogen, of course). An electron can, and will, slip through and among the atoms composing a metal wire where an atom cannot. (The ability of the electrons to pass through metals depends on the electronic structure of the metallic atoms themselves. The atoms of some substances, such as sulfur, glass, or rubber, have electronic structures that will not permit the free passage of electrons.)

The electron pressure in both rods acts as a driving force that will send those small particles through the wire. Since the electron pressure is higher in the zinc, the electrons move away from the zinc rod with greater energy and overwhelm the feeble flood of electrons moving out of the copper. The net movement of electrons, then, is from a region of higher electron pressure to one of lower electron pressure, just as a stone will move from a region of high potential energy to one of low, and just as heat will flow from a region of high temperature to one of low. This difference in electron pressure, which acts as a driving force behind electron flow, is called "electric potential" or "electromotive force" and, in the latter case, is abbreviated E.M.F.

As electrons leave the zinc rod the electron pressure decreases so that more zinc atoms can give off electrons. These in turn move away and still more zinc atoms give off electrons. This continues indefinitely and the steady electron flow is the familiar "electric current."

As electrons flood from the zinc into the copper, they are picked up by the copper ions surrounding the copper rod. Those ions become copper atoms. What happens then is that, while the electrons flow, zinc atoms form zinc ions that enter the solution at one end, and copper ions become copper atoms and leave the solution at the other end. The total content of positive ions in the solution does not change, nor does the quantity of positive charge. To be sure, positive charge accumulates in the portion of the solution near the zinc and becomes deficient in the portion near the copper,

but the negatively charged sulfate ion can travel through the porous partition from the copper end to the zinc end to keep things in electrical balance.

The electron flow is a form of energy and can be used to perform work. This "electrical energy" arises at the expense of the decrease in chemical energy involved in passing from copper ion plus zinc to zinc ion plus copper.

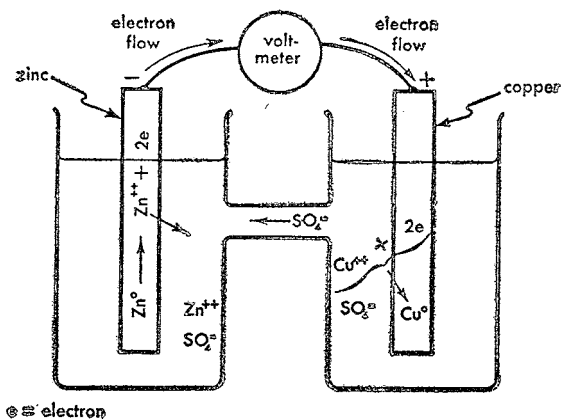


Figure 14. *The Flow of Electrons in a Voltaic Cell*

Such a device for converting the free energy of a chemical reaction into electrical energy is variously called a "chemical cell," an "electrical cell," an "electrochemical cell," or even a "voltaic cell" (see Fig. 14), after Alessandro Volta, the Italian scientist who first prepared a chemical cell in 1800. The particular cell involving zinc and copper is the "Daniell cell" named for the English chemist, John Frederic Daniell, who first devised one of this type in 1836.

(Notice that such cells were prepared and used a century before the electron was discovered. However, the use of electron flow, in the form of a branch of physics now called "electronics," has expanded unbelievably since the 1890s. The moral of this is that it is not necessary to understand the nature of a phenomenon to use it, but how it helps if we do understand it!)

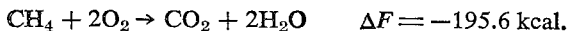
The greater the decrease in free energy involved in the chemical reaction that sets up the electron flow, the greater

the electric potential driving the electrons. It is possible to measure the electric potential easily enough (and it is measured as so many "volts" by means of a "voltmeter," again after Volta). From that potential, Gibbs's mathematics showed how the free energy decrease could be measured. In this way Gibbs tied not only chemical energy but also electrical energy into the thermodynamics that had been developed, originally, out of a consideration of heat engines.

This leaves some very important chemical reactions still unavailable for direct study. The combination of carbon and oxygen to form carbon dioxide has an equilibrium point so near the carbon dioxide extreme of the reaction that the free energy decrease associated with it cannot be calculated from that equilibrium point. Nor can an electrochemical cell, involving the combination of carbon and oxygen, be set up. (A cell of this sort, called a "fuel cell," would allow the production of electricity quite inexpensively and it would be a valuable and desirable device if it could be worked out.)

Fortunately, the equivalent of Hess's law can be used here. Hess's law, you remember, allowed chemical equations to be treated as though they were algebraic equations, with the ΔH values being manipulated analogously (see Chapter 8). The same can be done for ΔF values. By manipulating equations for which the free energy change can be measured directly one way or another, it is possible to come out with the value of the change for a reaction that cannot be measured directly.

Thus it is possible to say with as much certainty as anyone can reasonably expect in science that



and so on for many other reactions.

Chapter 11

THE ENCOURAGEMENT OF REACTION

The triumphant application of thermodynamics to chemistry in the 1870s and 1880s as described in the last couple of chapters is not without a flaw.

I have been saying over and over again that the combination of hydrogen and oxygen (and of carbon and oxygen) involves a large decrease in heat content and in free energy, and is spontaneous. Yet I have also mentioned several times, without special emphasis, that a mixture of hydrogen and oxygen must be heated before it will ignite; that it must be encouraged somehow to start its supposedly spontaneous combination. This is even truer in the case of the combination of carbon and oxygen. No householder worries that the coal in his cellar will combine with air before he is ready to put it in the furnace. In fact, it is rather difficult to start a coal fire.

Well, why should this be? If the free energy diagrams I drew in the previous chapter have any validity, then a mixture of hydrogen and oxygen ought to start reacting at once and without encouragement. In the mechanical analogy, a ball placed at the top of an inclined plane ought to roll downward at once when released, without requiring any push. Why, then, the need of a push in the form of heat in order to start a spontaneous chemical reaction? Is there something wrong with the principles of chemical thermodynamics?

One way out of the dilemma is to say that thermodynamic considerations predict the direction in which a chemical reaction will go, but not how quickly it will go. In other words, the hydrogen and oxygen in the mixture *are* combining; the carbon in the coal *is* combining with the oxygen of the air; both processes, however, are proceeding at so slow a rate that nothing can be observed even over long periods of time.

This may sound like refuge behind an uncheckable hypothesis, but the fact is that there is evidence that this is exactly so. Paper (made up of complex molecules containing carbon, hydrogen, and oxygen atoms) will combine with oxygen in a reaction involving a decrease in free energy. This is evident every time paper burns, releasing light and heat, in what seems quite evidently a spontaneous reaction. Yet paper doesn't combine with oxygen unless it is first ignited; at least as far as we can see. But is that so? Can we not see that it *does* combine even without ignition if we look far enough?

After all, the pages of an old book are yellow and brittle. The edges of the pages, which have been most nearly in contact with air, are yellower and more brittle than the interior of the page. The paper, in fact, looks slightly scorched. And it *is* faintly scorched; over the decades it has been very slowly combining with oxygen; very slowly "burning."

Take another example that is far more spectacular. Oily substances, with molecules made up largely of carbon and hydrogen, will combine with oxygen in a reaction involving a decrease in free energy and, naturally, the production of heat. Without ignition, such a reaction proceeds too slowly to be detected.

And yet the oily substances themselves are poor conductors of heat. If a quantity of oily rags are stored where air circulation is poor, the heat that develops with microscopic slowness, as the result of the very slow non-ignited combination of oil and oxygen, is neither conducted away nor blown away. Slowly, very slowly, it accumulates. As it accumulates and the temperature rises, the rate at which the oil combines with the oxygen also increases. Heat evolves at a continually increasing rate, and temperature rises until the rags actually ignite. The rags burn; the house may well burn with it; and the firemen say: "Spontaneous combustion," and warn the public about the proper disposal of oily rags.

So it would seem that spontaneous reactions are indeed spontaneous. What we need to explain their apparent reluctance to get started is not an overhaul of chemical thermodynamics but some principle that will arise out of a closer study of the rates of reaction ("chemical kinetics"). What makes a reaction go with undetectable slowness under some conditions and with explosive rapidity under others?

The first generalization to be discovered in connection with chemical kinetics is that a rise in temperature hastens all chemical reactions. This, of course, is part of the common experience of man, and even in prehistoric times men were aware of the fact that a good hot fire made for speed in cooking, metallurgy, glassmaking, and so on and that, on a lower scale, fruit juice would ferment more rapidly in a warm room than in a snowbank. It is a long road, however, from a general awareness of a principle to its accurate measurement under controlled conditions and only the latter really counts for much in science.

In the 1880s a number of chemical reactions were conducted at varying temperatures by various chemists and the rates at which they proceeded were carefully measured. It turned out that the rate just about doubled for every rise of 10°C. in temperature.

This seems reasonable in terms of kinetic theory. A chemical reaction between two substances cannot very well take place except when molecules of the two substances collide. As the temperature rises, molecular motion is hastened (see Chapter 6) and more collisions take place per second, hence the reaction is hastened.

But wait, let's think about that a little more. The average motion of molecules increases as the square root of the absolute temperature. In going from 20°C. to 30°C. , we are going from 293°K. to 303°K. , a rise in absolute temperature, and hence of molecular speeds, of only a few per cent. But if molecular motion increases by so little why should the reaction rate double; that is, increase by 100 per cent?

Furthermore, even if we forget that problem, another rises. If a mixture of hydrogen and oxygen is heated very slowly, nothing much seems to happen as the temperature rises, until about 600°C. is reached. Then there is a violent explosion!

This seems strange. If the rate of a reaction about doubles with every increase in 10°C. , one might expect the hydrogen and oxygen to combine at a smoothly increasing rate. This might very well get violently fast at a high temperature as observed, but ought to show intermediate rates in all gradations at lower temperatures. Instead, the rate increases from "too slow to measure" to "too fast to measure" over a very narrow temperature increase. Why?

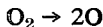
We had better take another look at what is involved in a chemical reaction and, for the purpose, let's keep to our mixture of hydrogen and oxygen.

So far, in talking about the combination of hydrogen and oxygen, I have ignored the details of the combination. I have said only that we begin with a mixture of hydrogen and oxygen at one free energy level, and end with water at a considerably lower free energy level. So far, so good; but what happens in the course of the combination? Surely hydrogen and oxygen molecules do not become water molecules by instantaneous abracadabra. The change must take place in definite time-consuming stages.

Consider that the hydrogen with which we begin is made up of molecules consisting of two hydrogen atoms held together by a chemical bond. Let's represent this as HH . The oxygen is similarly made up of molecules containing two oxygen atoms held together by a chemical bond (OO). The water with which we end is known, through many lines of evidence, to be made up of molecules containing an oxygen atom between two hydrogen atoms (HOH).

But in order to put an oxygen atom between two hydrogen atoms, the hydrogen atoms, previously bound to each other, must first be pulled apart. Similarly, in order to put a single oxygen atom between the two hydrogen atoms, that one oxygen atom must be pulled loose from its partner in the oxygen molecule.

Before ever we can form a water molecule, then, it would seem that the following reactions must take place:



But such decomposition of molecules into free atoms involves a large *increase* in free energy, as I pointed out in Chapter 7.

If we look at the combination of hydrogen and oxygen in this fashion, we see that it consists really of two reactions. The first is the conversion of molecules of hydrogen and oxygen to free atoms of those elements, while the second is the combination of the free atoms to form water. The free energy relationships might be expressed as shown in Fig. 15.

If, in Fig. 15, we look at the free energy of the molecular

mixture of hydrogen and oxygen (A) and compare it with that of water (B), then the free energy is downhill all the way, as shown by the dotted line. However, by introducing the intermediate step of atomic hydrogen and oxygen, we have a free energy value (X) far higher than either A or B. No wonder there is difficulty in having hydrogen and oxygen combine. It is one thing (using the mechanical energy analogy) to expect a ball to roll down a smooth inclined plane

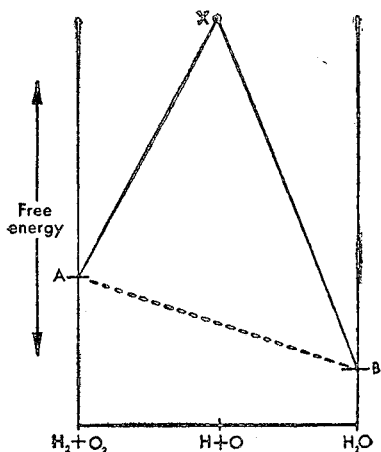


Figure 15. *Hydrogen-Oxygen Combination in Two Stages*

from A to B; it is quite another to expect it to follow the solid line and roll upward to the top of a mountain on the way.

In fact, now the reverse question arises. Why do hydrogen and oxygen combine at all? In Chapter 7, I said that even at a temperature of 3500°C . less than half the hydrogen and oxygen molecules were broken down to the atomic form—yet the gas mixture reacts with explosive violence even at 600°C .

The answer to this is that we must not assume that *all* the hydrogen and oxygen molecules must break down to atoms before combination to water can proceed. The production of even a few of them is sufficient to get things started at least.

The free energy diagram should not, therefore, have the high sharp peak, since the point X is only to be found in a mixture that consists entirely of hydrogen and oxygen atoms. Molecular hydrogen and oxygen, under ordinary conditions, never pass through such an all-atom stage in forming water. At best, only a small fraction of the system consists of the free atoms and so the intermediate rise in free energy is much gentler. The free energy diagram (see Fig. 16) contains a hump rather than a mountain peak. And yet, to be sure, our rolling ball finds the hump a problem too. It must

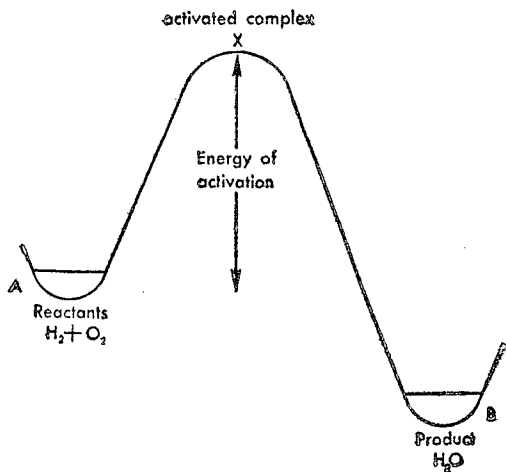


Figure 16. *The Energy of Activation*

be given a push to enable it to surmount the hump before it can roll down the rest of the way on its own. And the same is true, in energy terms, of a chemical reaction.

In 1889, the Swedish chemist Svante August Arrhenius worked out the mathematical relationships involved in this point of view and made the existence of the free energy hump clear not only in the combination of hydrogen and oxygen but in chemical reactions generally. The energy that must be put into a system to raise the free energy to the height of the hump in order that the reaction may then proceed spontaneously is now called the "energy of activation."

The easiest way to add energy equivalent to the amount necessary to "overcome the hump" and let the reaction proceed is to use heat. If a mixture of hydrogen and oxygen is sufficiently heated, it explodes; if paper or oily rags are sufficiently heated, they burst into flame. Thereafter, the reaction proceeds spontaneously with no further addition of heat necessary.

Looked at in this way, however, chemical reaction might seem an all-or-nothing proposition. It would seem that as long as the energy content of a system were insufficient to overcome the hump no reaction at all would take place. Then, as heat is added, a point is reached where, suddenly, the reaction goes. This point of view looks reasonable, if we think of hydrogen and oxygen doing nothing at one temperature and exploding at a slightly higher one; or of paper and oily rags inert and harmless at one temperature and in sudden wild flame at a slightly higher one.

But this is too simple a view. After all, I explained at the beginning of the chapter that paper and oily rags *do* combine with oxygen even at room temperature, although very slowly, and that so, presumably, do hydrogen and oxygen. Yet at room temperature the energy content of the system is far too low to meet the requirements of the energy of activation. Why does any combination at all take place then?

Once more, let's look at the hydrogen and oxygen mixture, assuming it to be at room temperature, about 25°C .

The Maxwell-Boltzmann equations (which I mentioned in Chapter 6) can be used to calculate the most probable speed and, therefore, the most probable kinetic energy of the gas molecules at that temperature. If this is done it is found that this energy is far less than that required to disrupt the bonds between the atoms of the hydrogen or oxygen molecules. Two molecules, with the typical energies to be expected at 25°C ., simply bounce apart after collision. One may gain energy at the expense of the other, but both remain intact.

However, the crucial point made by the Maxwell-Boltzmann equations is that the most probable energy of the molecules is only most probable; it is not universal. There is a wide range of energies at any temperature. While most molecules have energies somewhere near the average, there

are a few that have considerably more than average. There are even a very few molecules which, by fortuitous gains of speed during collisions, are zipping along with enough energy to disrupt themselves on collision or, perhaps, to disrupt the molecules with which they collide. Such molecules do indeed have energy equal to or greater than the energy of activation. The free atoms resulting from these very few energetic collisions can take part in reactions (either among themselves, or, more likely, with molecules that are still intact) to form water.

Even at room temperature, then, some hydrogen and oxygen molecules are energetic enough to make their way over the energy of activation humps and combine to form water. The percentage of molecules doing so is so small that the rate of water formation remains undetectably slow.

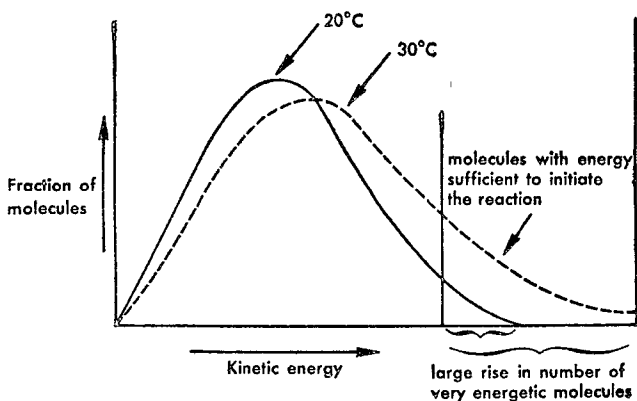


Figure 17. *Temperature Rise and Reaction Rate*

As the absolute temperature goes up, however, the average energy of the molecules goes up in proportion. A rise in temperature from 20° C. to 30° C. may represent, as I explained at the beginning of the chapter, an increase of only a few per cent in the most probable speed, but the number of molecules exhibiting very high energies doubles or even triples in quantity (see Fig. 17). (This is analogous to the manner in which a hot summer may have an average

temperature only 1° C. higher than that of a mild summer, but may nevertheless have two or three times as many days on which the temperature tops 90° F.)

Since the rate of reaction depends not on the average energy of the molecules but on the number of molecules with energy high enough to top the energy of activation, the rate doubles or more with each ten-degree (C.) rise in temperature.

In the case of the combination of hydrogen and oxygen, the rate of reaction is so terribly small at room temperature that even doubling the rate several times over, by raising the temperature, still leaves it too small to detect. Many reactions, however, have lower energies of activation than does the hydrogen/oxygen reaction. In the case of these others, quite a significant proportion of the molecules at room temperature possess energy enough to cross the relatively low hump, and the reaction rate can be easily measured. It was such low-hump reactions that gave the chemists of the 1880s the data they needed to measure the increase of rate with temperature.

But though this explains the doubling of reaction rates with small rises in temperature, there is still the matter of explosion to consider; the sudden transition from "too small to measure" to "too fast to measure."

This is a puzzle only if we assume that the only source of heat is what we are adding from outside the system. It is not. The formation of water molecules through the action of the free atoms produced is exothermic and liberates heat. The amount of heat liberated is very small and the increase in temperature resulting from it is very small. The gas mixture radiates the excess heat away and remains at constant temperature.

As the temperature rises due to heat poured in from the outside, the number of combinations increases and the total amount of heat produced by such combinations increases also. At a certain point the amount of heat produced each second is just a little higher than the amount of heat that the gas mixture can radiate off to its surroundings each second and the temperature does begin to rise. But as the temperature rises the amount of heat produced per second by way of internal combinations also rises and the surplus above the amount that can be radiated away is even higher, so that the temperature rises even more steeply.

It is a vicious cycle. Once the imbalance is set up, it brings about a consequence that strengthens the imbalance that sharpens the consequence that multiplies the imbalance that vastly enlarges the consequence. . . . At a certain temperature the mixture of hydrogen and oxygen may seem quite quiescent; then comes the boundary line beyond which radiation can't keep up and the imbalance starts and the result is so rapid that there is no time even to shut off the external heat source.

The same principle, in a much quieter manner, accounts for the burning of paper. Suppose you apply a match flame to one corner of a sheet of paper. The energy of activation at that corner is supplied and the molecules in the paper combine rapidly with oxygen. In doing so, they produce heat. This heat is sufficient to supply the energy of activation to neighboring molecules, which in turn produce heat and so on. The match flame may be withdrawn, for it is no longer needed. The paper burns of itself and an indefinite quantity of it will proceed to do so. As you well know, one smoldering match is quite enough to supply the initial energy that ends in burning down several square miles of forest.

The energy of activation of a particular reaction can be met by the addition of energy in forms other than heat. For instance, a mixture of equal volumes of hydrogen and chlorine will explode, if heated, more readily than will mixtures of hydrogen and oxygen, but will remain inert at room temperature, *if* kept in the dark. When exposed to light, particularly to the direct light of the sun or of a magnesium flare, a mixture of hydrogen and chlorine will explode even at room temperature.

This effect of light on a mixture of hydrogen and chlorine was first noted in 1801 by the English chemist William Cruikshank, and throughout the nineteenth century sporadic investigations of the phenomenon took place. A good understanding, however, required an insight into the nature of light that was beyond nineteenth-century physics and this will require a small detour.

The first modern investigation of light was that of Isaac Newton in the 1660s, who showed that sunlight could be broken up into a band of colored light (a "spectrum") by means of a triangular prism. Newton held that light consisted of a stream of exceedingly minute particles, hurtling

outward in all directions from its source at a tremendous velocity. (The actual velocity of light was first determined, with reasonable accuracy, in 1671, by the Danish astronomer Olaus Roemer. The accepted value today is 186,282 miles per second—certainly tremendous enough.)

A contemporary of Newton, the Dutch astronomer Christian Huyghens, offered an alternative interpretation, to the effect that light consisted of a wave motion.

During the 1700s, Newton's "corpuscular theory" was the more popular. However, a series of investigations in the first half of the nineteenth century seemed to show conclusively that light consisted of waves after all and the "wave theory" was firmly established.

Each individual wave of light is extremely short, with a length that varies from color to color. The exact "wave length" can be calculated in a number of ways, and every line of investigation agrees in making the wave length of red light fall in the range of 0.0000007 meters.

Obviously the meter is too long a unit with which to measure light waves. A "micron" is a millionth (0.000001) of a meter and a "millimicron" is a thousandth of a micron or a billionth (0.000000001) of a meter. The Swedish astronomer Anders Jonas Ångström, who carried on researches on light in the mid-nineteenth century, worked with a unit of length equal to a tenth of a millimicron or a ten billionth (0.0000000001) of a meter. That has been named an "Ångström unit" in his honor and this is usually abbreviated as "Å." The wave length of red light can therefore be expressed as in the neighborhood of 7000 Å (see Fig. 18).

The wave length of light varies continuously, its color changing by imperceptible degrees. The color is distinctly orange in the 6000 Å range, yellow by about 5700 Å, green by 5300 Å, blue by 4800 Å, and violet by 4200 Å. These are of course average values. Each color fades insensibly into the next on either side. Red fades into darkness at an extreme of 7600 Å, and violet at the other extreme of 4000 Å.

As important as wave length is the concept of "frequency," the number of waves that enter the eye (or an appropriate instrument) each second. Now, in one second light travels 186,282 miles or about 3,000,000,000,000,000 Ångström units. The number of times the very longest of the red waves will fit into that stretch is 3,000,000,000,000,-

000,000/7600 or about 400,000,000,000,000. This means that four hundred trillion waves of deep red light enter your eye every second. That is its frequency.

Naturally, the shorter the wave length, the greater the number of waves that fits into a one-second stretch of light. Therefore, the shorter the wave length, the greater the frequency. The frequency of the violet light of shortest wave length is about 730,000,000,000,000 or just about twice as high as that of red light.

About 1860, Maxwell set about studying the nature of electric and magnetic forces that made themselves felt throughout space in the neighborhood of a magnet or an electric charge. His reasoning showed electricity and magnetism to be inextricably related, producing a single "electromagnetic field" through space. Furthermore, he showed that any oscillating electric charge would radiate energy ("electromagnetic radiation") in wave form with a frequency just equal to the rate of oscillation. This energy would travel through space at the velocity of light.

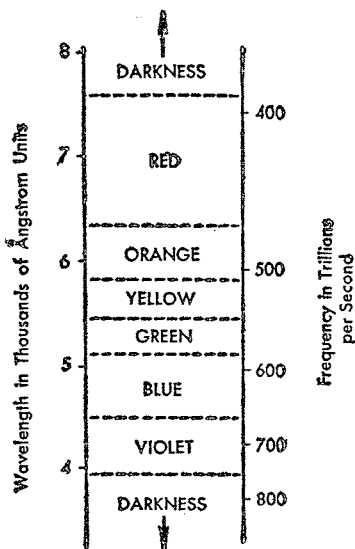


Figure 18. *The Visible Spectrum*

To Maxwell, the similarity between his theoretically deduced velocity of electromagnetic radiation and the known velocity of light seemed too much to ask of coincidence. He decided that light was a form of electromagnetic radiation.

Of course, one might ask what the charge was that oscillated with a frequency in the hundreds of trillions in order to produce light, and for that Maxwell had no answer. It was only with the discovery of the inner structure of the atom in the 1890s that an answer became possible. Then, the Dutch physicists Hendrik Antoon Lorentz and Pieter Zeeman were able to show that the oscillators were the electrically charged electrons that made up the outer regions of the atoms. Changes in the energy levels of these electrons produced light waves.

It seemed to Maxwell, further, that charges could oscillate with any frequency and that therefore there should be electromagnetic radiation of any frequency. In particular, there should be radiations with lower frequencies and longer wave lengths than red light; and also with higher frequencies and shorter wave lengths than violet light.

Now radiation beyond the red ("infrared") and violet ("ultraviolet") had already been discovered. The eye is not sensitive to such radiation, so they are invisible, but they are nevertheless real in every physical respect. About 1800 the German-born English astronomer William Herschel held a thermometer just beyond the red end of the spectrum, where, to the eye, nothing existed. The temperature rose even faster than if it were held in the visible portion of the spectrum, so radiant energy must exist there.

At about the same time a German physicist, Johann Wilhelm Ritter, found that while blue and violet light brought about the breakdown of silver nitrate to metallic silver, blackening the previously white compound, the invisible region of the spectrum just beyond the violet did so even more rapidly. Radiant energy existed there as well.

However, infrared and ultraviolet were neighbors of the visible spectrum and Maxwell's theory called for radiation to exist in regions far removed from the visible. Evidence of such extreme radiation was lacking in his lifetime.

Maxwell died in 1878, at the age of only forty-eight, and did not live to see his theory confirmed. In 1886, however, the German physicist Heinrich Rudolph Hertz set up a circuit in which he caused an electric current to oscillate at

quite low frequencies. This, by Maxwell's theory, should produce radiation at equally low frequencies (and very long wave lengths). Hertz was actually able to detect such radiation, and these are now known as "radio waves."

Radio waves have frequencies as low as 1,000,000 or less. Shortwave radio waves, or "microwaves," have frequencies up in the 10,000,000,000 (ten billion) range. Radio waves fade gradually into microwaves which fade gradually into infrared which fade gradually into the visible region. There are no gaps. Every frequency (and wave length) is represented.

What about the other end of the spectrum? In 1895 the German physicist Wilhelm Konrad Roentgen detected a new kind of radiation liberated by the impact of speeding electrons on matter. These "X rays," as he called them, have frequencies much higher than even the ultraviolet. The frequencies are in the millions of trillions. As eventually appeared, they were set up by changes in the energy levels of the inner electrons of the atom. A year later it was discovered that uranium gave off hitherto unknown radiations and, as it turned out, these included a variety of electromagnetic radiation, called "gamma rays." These were of frequencies even higher than that of the X rays, being in the billions of trillions. These were set up by changes in the energy levels of charged particles within the tiny atomic nucleus itself.

Again, gamma rays fade into X rays, which fade into ultraviolet, which fade into visible light. The entire "electromagnetic spectrum" (see Fig. 19) contains some 60 octaves all told (an "octave" being a range of frequencies over which the value of the frequency doubles). Of these only a single octave consisted of visible light. By 1900, Maxwell's theory stood wholly confirmed.

Light is a form of energy but, as became clear with the opening of the nineteenth century, the various colors of light are not equally energetic. When light brings about the breakdown of silver compounds to metallic silver (a phenomenon that is the basis of photography) it is the high-frequency light at the blue end of the spectrum that does so. The low-frequency light at the red end of the spectrum is without effect.

This and other lines of evidence made it appear that the

energy content of high-frequency light was greater than that of low-frequency light. Ultraviolet light was discovered because it was still more effective in decomposing silver compounds than was any form of visible light. Ultraviolet light could redden and burn the skin as well. X rays can penetrate matter deeply and gamma rays still more deeply, and both can bring about reactions in tissue that require a large input of energy (and that are harmful).

The greater energy of high-frequency radiation is also indicated by the behavior of heated objects. Some of the energy they gain with rising temperature is liberated in the form of radiation. The higher the temperature rises, the higher the frequency of the emitted radiation. Thus a piece of metal, when heated, will begin to radiate energy (and will feel warm or even hot to a hand placed nearby) while re-

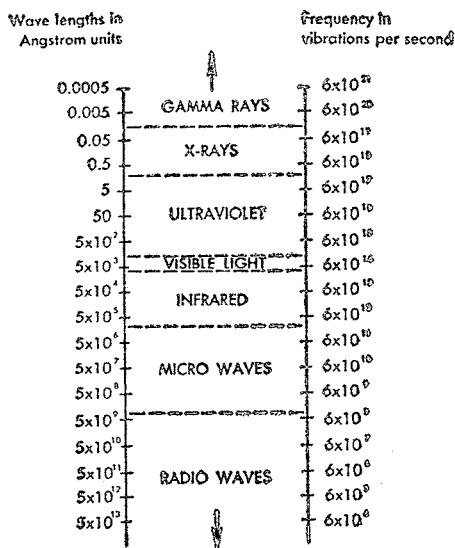


Figure 19. *The Electromagnetic Spectrum*

maining completely invisible when in the dark. It is radiating in the infrared. As the temperature rises, enough of the radiation is in the visible region of the spectrum to cause

the metal to glow a deep red. To be "red hot" in this fashion requires a temperature of about 650°C .

At still higher temperatures, the radiation continues to increase in frequency, so that the metal will glow cherry-red and will be a distinct "yellow hot" at 2000°C . The sun, with a surface temperature of 5700°C ., is "white hot" and radiates an appreciable amount of ultraviolet. There are even a very small percentage of stars with far higher temperatures that radiate chiefly in the blue end of the spectrum and the ultraviolet. The sun's corona, with a temperature of a million degrees and more, actually radiates in the X ray region.

Interest in the exact details of the manner in which radiation was emitted at various temperatures grew toward the end of the nineteenth century. The matter was simplified by assuming the existence of a body that would completely absorb all radiation, of whatever frequency, falling upon it. Such a body would reflect no radiation and would therefore be a "black body." If such a black body were heated it should radiate all frequencies with equal ease. (A body that is not "black" and that absorbs only certain frequencies would radiate those frequencies only when heated.)

But if a black body could radiate all frequencies with equal ease, then there was a paradox to be explained away. There were many more high frequencies of radiation than low frequencies (just as there are many more numbers over a trillion than under a trillion). For that reason, if light "chooses" from among all possible frequencies equally, virtually all of it would be radiated at the very numerous high frequencies and very little at the relatively few low frequencies. However, when a black body was approximated and heated, no such thing happened. Even at quite high temperatures, most of its radiation was in the visible spectrum and in the still lower frequencies. It took very high temperatures to produce an appreciable amount of high-frequency radiation.

No explanation of this was available until 1900, when the German physicist Max Karl Planck had an idea. What if radiation, like matter, was discontinuous? Just as matter existed in small bits called atoms, why couldn't radiation exist in small bits too? To explain the fact that high-frequency radiation contained more energy than low-frequency radia-

tion, it was only necessary to suppose that high-frequency radiation occurred in larger bits.

Planck called the bits of energy "quanta" (singular, "quantum," meaning "how much?" in Latin) and proposed that the energy content of a single quantum was proportional to the frequency of the radiation. If we let energy be represented by e and the frequency of a particular type of radiation by the Greek letter ν (or "nu"), then Planck was suggesting that:

$$e = h\nu$$

where the proportionality constant, h , is now universally known as "Planck's constant."

Now if a body is heated, it could not be expected to emit radiation in all frequencies with equal ease, even if it were a black body. Since gamma rays have about a billion times the frequency of infrared rays, it would take a billion times as much energy to form a gamma-ray quantum as to form an infrared quantum. At low temperatures, with little energy available, only infrared quanta could be formed with any ease at all. As temperature rose and the available energy increased, larger and larger quanta of higher and higher frequencies of light could be formed. In fact, by use of the "quantum theory" the manner of radiation of a black body could be described exactly; something that could not be done before.

The theory was so revolutionary, however, that for a while few physicists would go along with it. Then in 1905 the German-born physicist Albert Einstein used the quantum theory to explain the manner in which light caused certain metals to expel electrons.

It had been observed toward the end of the nineteenth century that certain metals emitted electrons when exposed to light. The energy of the electrons emitted did not depend on the intensity of the light but on its frequency. For instance red light might not cause the emission of any electrons at all, no matter how intense it was. Yellow light might bring about the emission of very low-energy electrons. Intensifying the yellow light might cause the emission of more electrons but they would all be very low-energy. Blue light would bring about the emission of energetic electrons. Very

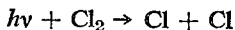
feeble blue light might bring about the emission of very few electrons, but those that were cast off were high-energy.

Einstein pointed out that this could be explained if it were to be assumed that energy could be absorbed only in whole quanta. A small quantum of red light would have insufficient energy to eject an electron and increasing its intensity would have no effect, for an atom would always get rid of one quantum before absorbing the next. The larger quanta of yellow light absorbed all at once would kick out the electron with a small quantity of energy, and the still larger quanta of blue light would kick it out with a larger amount of energy.

In this way Planck's quantum theory was established and gained acceptance among scientists.

Now we can go back to the mixture of hydrogen and chlorine and see, with new insight, what it is that light does to it. When a gas mixture is simply heated over a hot plate or over the flame of a Bunsen burner, the radiated heat entering the mixture is mostly in the infrared, with a little in the low-frequency portion of the visible range in the case of the flame. The quanta involved are small. It is not until the gas molecules have absorbed enough quanta to gain considerable kinetic energy that collisions will be strenuous enough to disrupt them.

The quanta of the comparatively high-frequency violet and ultraviolet radiation in sunlight or in a magnesium flare are much larger. A molecule of chlorine (composed of a pair of atoms less firmly bound than is the case of the atoms in either the hydrogen or oxygen molecule), absorbing such a quantum, gains all at once enough energy for disruption. If we let $h\nu$ represent the quantum, we can write the equation:

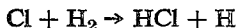


Each atom of chlorine thus formed can then react with a hydrogen molecule to form hydrogen chloride.

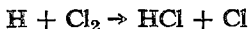
If this were, however, the whole story, then a single quantum would be responsible for the formation of two molecules of hydrogen chloride and the total number of quanta required to convert all the gas mixture to hydrogen chloride would be equal to the number of chlorine molecules present.

Actually, though, far less in the way of radiation is required. Careful measurement of the amount of radiation supplied and the amount of hydrogen chloride produced would indicate that a single quantum can be responsible for the formation of anywhere from 10,000 to 1,000,000 molecules of hydrogen chloride.

Since the quantum simply lacks the energy to perform that large a task by itself, something else must be involved. The explanation currently accepted was first proposed in 1918 by the German physical chemist Walther Hermann Nernst. He suggested that the disruption of the chlorine molecule by a high-energy quantum was only the first step. The chlorine atom produced reacts with the hydrogen molecule to produce not only a molecule of hydrogen chloride but also a free hydrogen atom, thus:



The free hydrogen atom produced could, in turn, react with a chlorine molecule to produce a new atom of free chlorine:



The chlorine atom would then react with another hydrogen molecule to produce a hydrogen atom that would react with another chlorine molecule and so on indefinitely—with a molecule of hydrogen chloride being produced at each step.

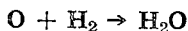
The function of the quantum of energy is merely to initiate such a series of reactions; a "chain reaction," as it is called.

The question arises as to why the chain reaction does not continue on until every molecule of hydrogen and chlorine in the mixture is consumed and converted to hydrogen chloride. The chain, however, is not immortal. Sooner or later some reaction will take place with some impurity present in the mixture, with a water molecule or an oxygen molecule. Such a reaction may not produce a free atom to replace the one consumed and a chain reaction comes to an end.

Even in the absence of all impurities (and the removal of enough impurities to keep the chain long-lived is not as simple as one might suppose) there is still the chance that

a free atom may come to an end by reacting with some molecule in the wall of the container. There is even the chance that one free atom may collide with another, forming a molecule and bringing two chains to an end. Even under ideal conditions, then, a chain reaction cannot go on forever.

Naturally a gas mixture that offers an overwhelming chance for ending any chain that starts up is not likely to be stirred by light radiation to any visible extent. For instance, a mixture of hydrogen and oxygen is not affected by light radiation particularly. Even if a quantum of radiation large enough to disrupt the oxygen molecule is absorbed, the free oxygen atom produced need only collide with one of the hydrogen molecules to end the chain before it starts:



I have gone into some detail on the nature of light and its effect on chemical reactions because at the end of the book we shall come back to the subject in a supremely important aspect. However, the path we are painstakingly following leads us away from that final subject now and toward many another curious item instead.

Chapter 12

ENCOURAGEMENT WITHOUT ENERGY

Let's stop for a moment now and take our bearings. We began by searching for a distinction between the living and the non-living and that led us to a consideration of the thermodynamics of heat engines.

However, we decided that living organisms were not heat engines but that there was a possibility they obtained their energy by processes analogous to combustion. This led to a consideration of chemical energy and the manner in which that fitted in with the laws of thermodynamics. This, in turn,

led to the problem of why just those reactions involved in combustion and suspected of involvement with life, although labeled spontaneous by the laws of thermodynamics, nevertheless did not appear spontaneous.

That brought us to the subject matter of the previous chapter—the energy of activation that had to be supplied to get a spontaneous reaction “over the hump.”

How does this help us, now, toward our consideration of the nature of life? Suppose that reactions similar to those involved in combustion go on in the human body. Presumably, they possess an energy of activation that the body must supply before those reactions can proceed spontaneously and supply the body in turn with the energy it needs. And yet—nothing in the previous chapter seems to be applicable to living tissue.

I have described two ways in which the energy of activation can be supplied, through heat and through light, and surely the body uses neither. Living tissue never rises above 37° C. or so in temperature and this simply does not represent enough heat to supply the energy of activation for the reactions involved in combustion. And living tissue, though it is, on occasion, a source of light (as in fireflies) is never a source of the type of high-energy radiation that has a chance of meeting the energy of activation needs. How, then, does the living organism deal with the energy of activation?

Is it possible that here, finally, is a clear example of a way in which living tissue disobeys the laws of thermodynamics? May we even wonder if this is not itself the long-sought criterion that separates life from non-life?

Before turning to this particular solution of the dilemma, however, let us see if anywhere in the inanimate world there are phenomena that can make a spontaneous reaction proceed *without* supplying the energy of activation that may ordinarily be required. If so, then something of the sort may go on in living tissue.

The first important case of such a phenomenon was discovered in connection with the manufacture of sulfuric acid.

Sulfuric acid was discovered, along with several other strong acids, in the Middle Ages as one of the important, but disregarded, products of alchemical research. (The alchemists, themselves, only too often interested themselves

uselessly in the production of gold and attached too little importance to the real fruits of their labors.)

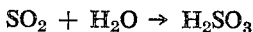
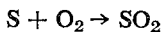
Sulfuric acid is many times stronger than the strongest acid (the acetic acid of vinegar) known to the ancients. It could bring about a number of chemical reactions that weaker acids could duplicate only slowly if at all. The strong acids generally, and sulfuric acid in particular, represented, then, a strong weapon in the chemical armory, in connection with both pure research and industrial processes. In fact, even today, aside from ubiquitous air, water and, perhaps, salt, sulfuric acid is the chemical used most frequently and in the greatest quantity by industry. Some fifteen million tons of it are manufactured each year and it has been said that the extent of industrialization of a nation can be measured by the amount of sulfuric acid it consumes.

It follows, then, that the development of techniques for the production of sulfuric acid cheaply and in quantity was of the highest importance. At the dawn of modern times, however, production was difficult and the supply limited.

This was not because of any difficulty in obtaining source material. The ultimate source was sulfur (S), a substance known even to the ancients, common enough, and for centuries obtained in reasonable quantity and with reasonable ease, in such places as Sicily.

Sulfur will burn easily, combining with oxygen to form a gas with a choking odor called "sulfur dioxide" (SO_2). It is this gas that is responsible for the familiar "smell of sulfur," and not sulfur itself, which is odorless.

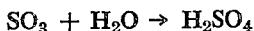
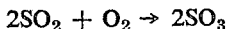
Sulfur dioxide can be dissolved in water. It thereupon combines with the water molecule to form "sulfurous acid" (H_2SO_3).



Sulfurous acid is only a moderately strong acid and, moreover, is not very stable. It is not the desired sulfuric acid, which contains an additional atom of oxygen and has the formula H_2SO_4 .

To add on the extra oxygen atom is difficult. To be sure, sulfur dioxide can be further combined with oxygen to form

"sulfur trioxide" (SO_3), and sulfur trioxide on solution in water will form the necessary sulfuric acid.



The trouble with this procedure, which seems so straightforward on paper, is that although the combination of sulfur dioxide and oxygen represents a decrease in free energy and is spontaneous, the energy of activation is so high that only small quantities of sulfur trioxide can be formed in this manner through direct combination of sulfur dioxide and oxygen. Through the seventeenth and early eighteenth centuries, then, sulfuric acid could be produced only in dribblets and could be used only sparingly. It took a week of hard work to produce two pounds of sulfuric acid.

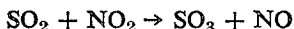
Then, in the 1740s, a manufacturer of sulfuric acid, Joshua Ward, found that he could produce much more sulfuric acid out of the same batch of burning sulfur by adding saltpeter (potassium nitrate) to the sulfur. Whether this was done accidentally to begin with or in an effort to make the sulfur burn more briskly by making it more closely resemble gunpowder (which consists of sulfur and saltpeter, plus powdered carbon), I don't know. In any case, he patented the process and the price of sulfuric acid dropped drastically.

With the production of sulfuric acid increasing, the process was no longer carried on in fragile glass equipment but in much larger chambers made of lead. The process of burning sulfur to which nitrate has been added is called the "lead-chamber process."

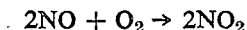
For half a century manufacturers continued adding nitrate without knowing why. No nitrogen-containing compound appeared in the product. In fact, at the conclusion of the process, the nitrate remained in the chamber and could be used over and over again. It supplied nothing to the reaction, at least in appearance, and it was not consumed. What then did it do?

In 1806 two French chemists, Charles Bernard Désormes and Nicolas Clément, advanced an explanation that contained a principle that is accepted to this day. When sulfur and nitrate burn, two gases are formed, sulfur dioxide (SO_2)

and nitrogen dioxide (NO_2). The nitrogen dioxide, they suggested, passed on one of its oxygen atoms to the sulfur dioxide, forming sulfur trioxide and leaving behind nitrogen oxide (NO) thus:



The nitrogen oxide then absorbed oxygen from the air to become nitrogen dioxide again:



In this way the nitrogen dioxide takes the role of a middleman, passing an oxygen atom to the sulfur dioxide, picking up another from the air which it passes to another molecule of sulfur dioxide, then picking up still another from the air which it passes on to still another molecule of sulfur dioxide and so on. For all this frantic activity, the nitrogen dioxide is not consumed and at the end of the reaction reforms nitrate and seems to be unchanged and to have been uninvolved.

Since the time of Désormes and Clément, the picture has been refined and complicated. It is now known that nitrogen dioxide and sulfur dioxide combine, along with water and oxygen, to form a complex "addition product" that breaks up to form sulfuric acid and restore the nitrogen dioxide. However, the principle remains the same—participation and regeneration.

From an energy standpoint, what has happened is that the high energy of activation involved in the direct combination of sulfur dioxide and oxygen has not been supplied; it has been bypassed (see Fig. 20). The direct combination was avoided altogether. Instead, sulfur dioxide combines with nitrogen dioxide (to form the addition product, by modern thinking) in a reaction that involves a much lower energy of activation; while the sulfuric acid is formed and the nitrogen dioxide regenerated by a second reaction that also involves a low energy of activation.

You are not getting something for nothing. The laws of thermodynamics are not defeated. The free energy of sulfur dioxide on one hand and of sulfur trioxide (or sulfuric acid) on the other remains the same. The over-all movement is downhill both ways. However, instead of passing from sulfur

dioxide to sulfur trioxide over a high hump, the route (thanks to the addition of nitrate) is directed over a pair of low

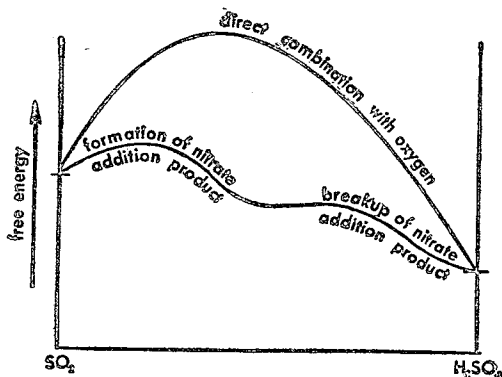


Figure 20. *Bypassing the Energy of Activation.*

humps. This interpretation, in terms of energy of activation, was first made clear for all phenomena of this type, in 1894, by the German physical chemist Wilhelm Ostwald.

If the sulfur dioxide reaction with oxygen to form sulfur trioxide is viewed as a railway journey from a valley into a lower valley over a high mountain pass, then the addition of nitrate is equivalent to the opening of a new route over lower mountain passes. If the trains are chronically short of fuel, very few will make their way up to the high pass and the passage from valley to valley will be rarely accomplished. The opening of the new route, even without an increase in fuel, will greatly increase traffic between the valleys.

As the nineteenth century progressed, more examples were found where small quantities of some substances hastened reactions involving large quantities of reacting material, without that small quantity of hastening agent being consumed. In 1812, for instance, a Russian chemist, Gottlieb Sigismund Kirchhoff, found that, if he boiled starch in water to which a bit of dilute sulfuric acid had been added, the starch was broken down and a sugar formed that was identical with that found in grapes (and later called "glucose").

Now if the acid were omitted and the starch merely boiled

in water, this did not happen. Furthermore, when the starch was broken down by the action of the acid, the acid was not consumed in the process. The final sugar solution was as acid as the original starch suspension had been.

Again, the English chemist Humphry Davy discovered in 1816 that wires of certain metals encouraged the combination of various vapors, such as that of alcohol, with oxygen. This was particularly true of the rare and inert metals, platinum and palladium.

This line of research led to the discovery by the German chemist Johann Wolfgang Döbereiner that platinum powder would bring about the combination of hydrogen and oxygen at room temperature. He even devised an automatic lighter (Döbereiner's lamp) years before the first practical matches were invented. His lamp consisted of a jar of acid so designed that the acid could be allowed to rise, at will, into an inner chamber that contained a zinc rod. The reaction between zinc and acid produced hydrogen, which emerged from an orifice at the top of the jar. The jet of hydrogen impinged on a container of platinum powder and promptly burst into flame.

Of course, platinum powder is expensive, and it quickly loses its unusual properties because of the presence of small quantities of impurities in the hydrogen formed. At best, then, Döbereiner's lamp can only be looked upon as an amusing chemical toy. However, it did dramatize the odd action of platinum. It also showed that the reactions involved in combustion (and, presumably, in the human body) could be made to go at room temperature without the necessity of adding the energy of activation (which is what we set out at the beginning of this chapter to discover).

In 1836 these and a number of other examples of such phenomena came to the attention of the Swedish chemist Jöns Jakob Berzelius. He, at this time, ruled the world of chemistry with an iron hand, for he interested himself, with fruitful results, in every branch of the quickly developing field.

Berzelius organized the scattered examples of the phenomenon, pointed out their importance, and suggested that the name "catalysis" be applied to it. This comes from Greek words meaning "to break down" and may have been suggested by the breaking down of starch to sugar in the presence of acid. The term was accepted and still is.

Any substance that, by its presence in small quantities (sometimes very small indeed), hastens the reaction of large quantities of other substances, without itself being consumed or substantially altered in the process, is a "catalyst." Acid is the catalyst breaking down starch to sugar. Platinum is the catalyst bringing about the combination of hydrogen and oxygen at room temperature. Nitrogen dioxide is the catalyst that brings about the production of sulfuric acid in the lead-chamber process.

The new discoveries, by the way, spelled the doom of the lead-chamber process that had started it all. In 1831 an English chemist, Peregrine Phillips, found that platinum could be used to catalyze the conversion of sulfur dioxide to sulfur trioxide more efficiently than nitrate could, and patented the process. This new "contact process," in which sulfur dioxide made contact with platinum in order to combine with oxygen, replaced the lead-chamber process by the turn of the century, after methods were discovered to prevent the expensive platinum from losing its catalytic properties too soon.

It is tempting to think of factors such as heat and light as catalysts, too, for a little bit of heat will hasten any reaction, while a little bit of light will explode a mixture of hydrogen and chlorine that had until then lain torpid. In fact, a process of the latter type is often spoken of as "light-catalyzed."

However, it is customary to apply the term "catalysis" and "catalyst" to material substances only.

It is rather easy to see that catalysts such as nitrogen dioxide or acid can actually take part in a reaction and be regenerated. They are active substances that take part in chemical reactions whenever given the chance. This is not so in the case of the most spectacular catalyst to be discovered in the early nineteenth century—platinum.

Platinum is an inert metal that combines with other substances only with the greatest reluctance and under the severest urging (to speak metaphorically). How, then, can it take part in a variety of reactions in order to bring about catalyzed combinations? That it does take part seems necessary, for to suppose that it merely exerts a mystical influence over other molecules without itself being involved is unthinkable.

In 1833 the English scientist Michael Faraday made a suggestion in this connection. He found that a clean piece of platinum foil would catalyze the combination of hydrogen and oxygen, but that a small quantity of carbon monoxide added to the gaseous mixture would suffice to end the catalysis. The carbon monoxide "poisoned" the platinum. (There is an interesting similarity here with respiration. A bit of carbon monoxide added to our own supply of oxygen will quickly poison *us*.)

Faraday suggested that the oxygen and hydrogen formed films on the platinum surface. By forming a tight film, the oxygen and hydrogen would be much closer on the surface than they would be in the original gaseous mixture. The proximity would be the equivalent, in an open gaseous mixture, of what could only be reached by great pressures, and it was known that increasing pressure on a gas mixture would often hasten the combination. He then pointed out that carbon monoxide, if present even in small quantities, might form a film on the surface preferentially, occupying it so that neither hydrogen nor oxygen could get at it, thus ending the catalysis.

This belief that it is the surface of some catalysts that counts ("surface catalysis") still prevails, albeit with refinements. The mere existence of a film of gas on such a surface is not apparently the answer, for the high pressure of which it is the equivalent cannot, on the basis of later information than was available to Faraday, sufficiently hasten the reaction.

In 1916 the American chemist Irving Langmuir modified the Faraday theory by suggesting that the film of gases was held by actual chemical bonds to the platinum atoms at the surface.

The platinum atoms at the surface are not completely surrounded by other platinum atoms and so their bonds are not entirely taken up. The bonds facing the open air are available for acting as an attractive force between the platinum (Pt) atoms and other types. The superior catalytic activity of platinum powder over solid platinum could then be attributed to the fact that in platinum powder there is a much larger surface per weight of platinum, and that a much larger fraction of the platinum atoms are at the myriad surfaces of the myriad particles.

Thus platinum might be viewed as entering into a reac-

tion with oxygen after all. All the surface platinum atoms, by hooking onto oxygen atoms, form a kind of compound with oxygen, a "platinum oxide." The energy of activation for the formation of this oxide is very low, and the combination of hydrogen with the oxygen of this oxide (as opposed to the free molecules of oxygen in the body of the gas mixture) is also low. The combination then proceeds quickly, for the kinetic energy of the gas molecules at room temperature is more than sufficient to supply the necessary energy of activation.

With the combination complete, the platinum atom finds itself attached to a molecule of water. The bond between platinum atoms and water molecules is quite weak, however, and the pounding of the gas molecules pushes it away, regenerating the free platinum atom and leaving it clear for the binding of another oxygen atom.

The poisoning of a platinum surface by carbon monoxide comes about through the ability of the platinum atom to bind a carbon monoxide molecule as well as oxygen molecules. The carbon monoxide molecule, once bound, does not react with hydrogen (or with anything else in the system) to form another molecule that is weakly held. It therefore persists in its position. Every platinum atom occupied by a carbon monoxide molecule remains occupied; more and more of the platinum surface is taken up and before long the catalytic power with respect to hydrogen and oxygen is gone.

Even surface catalysis does not alter the fundamental energy relationships in a reaction. It may reroute the pathway of reaction to lead it over lower energies of activation and hasten the over-all rate, but catalysis does not and cannot alter the over-all free energy difference of a reaction.

This means that a catalyst cannot and does not reverse the direction of a spontaneous reaction. If hydrogen spontaneously combines with oxygen to form water, a catalyst may hasten the process by lowering the energy of activation. No catalyst, however, can bring about the massive dissociation of water to hydrogen and oxygen with nothing more than the molecular motion at room temperature as the source of energy.

Furthermore, in any reversible reaction the position of the equilibrium point (which you will remember is at the point of minimum free energy) is not altered by the presence of a catalyst. The catalyst may (and does) hasten the reaction

in both directions and succeed in causing the equilibrium point to be reached more rapidly. The catalyst cannot, however, alter the position of that equilibrium point by one iota.

With this consideration of catalysis behind us, I think it is time to turn—equipped, as we are, with some basic concepts—to living tissue at last. In the second part of this book I hope to show you to what extent these concepts, derived from and based upon the phenomena of the inanimate universe, can be applied to the subtle activity animating the human body.

Chapter 13

LIFE AND NON-LIFE AGAIN

All the findings and deductions concerning the conservation of energy and the increase in entropy, concerning free energy and catalysis, were obtained from a consideration of the inanimate world. My intention, in spending the first half of this book in describing and explaining these findings and deductions, has been to use them as a means of helping to explain the inner workings of living tissue.

But this involves the assumption that generalizations based on inanimate phenomena can be applied to living things. Is such an assumption valid? In other words, do the same laws of nature govern the behavior of a rock and of the lizard upon it?

Through most of man's history such a question would have seemed foolish indeed and the answer would have been given as a clear "No!" Whereas I have been spending many pages in a careful and laborious effort to prepare the groundwork for a distinction between life and non-life, thus implying that it is difficult to make such a distinction—such difficulty was not even suspected until about a century ago. Rather it was always believed till then that life was fundamentally, irrevocably, and *obviously* distinguished from non-life. What fool could doubt for a moment, or fail to see, the fundamental cleavage of behavior and nature between the rock and the lizard!

In fact, it seems rather natural to divide the objects of the universe, without confusion or ambiguity, into three great classes: the inanimate world consisting of sea and land, of rocks and sand and air; the plant world, consisting of ob-

jects that are born and live and die, but are rooted to the ground and do not move independently; and the animal world, consisting of objects that are born and live and die and are, in addition, possessed of the faculty of willful motion. We take just this view when we play the game of Twenty Questions and begin by trying to determine whether the object being thought of is "animal, vegetable, or mineral."

Further thought, however, would make the distinction between plants and animals far less distinct than that between either and the mineral kingdom. For one thing, man could utilize as food members of both the animal and vegetable kingdom. For another, the coming of the microscope in the seventeenth century and the discovery of forms of life too small to be seen with the unaided eye further blurred the distinction, for many microorganisms could not be clearly stated to belong to either kingdom.

A broader division seemed natural then: the animate (including both plants and animals) on one hand, and the inanimate on the other.

Here, surely, one had two sets of objects of completely different nature and therefore, it was almost inevitable to assume, two different sets of laws of nature. Well into the nineteenth century the distinction between life and non-life and the existence of a double set of laws was taken for granted.

This held, not only for the discrete objects themselves, the rock and the lizard, but even for the chemical substances making up the objects.

Crush the rock and the powder is stable and inert. It does not change with the years. Mix it with water, then drive the water off by heating, and the powder retains its identity. Heat the powder till it glows, cool it, and it remains essentially unchanged.

On the other hand, mince the lizard, and you have a mass of soft bloody meat that will in a short time corrupt, raise a stench, and feed maggots. Heat the meat and it will sizzle and char and change beyond recognition. Nor will it return to its original state when cooled again.

This radical difference in behavior persists when you work with a pure chemical isolated from the general mass. From a rock, it may be possible to isolate pure aluminum silicate or pure copper. From a living organism, it may be possible

to isolate a pure sugar or a pure oil. The aluminum silicate or copper can be heated, then cooled, without essential change. If the sugar is heated, on the other hand, it will char and no longer be sugar, while the oil will burn and no longer be oil.

This distinction between two sets of substances forced itself on chemists in the very dawn of the science. In the seventeenth century, for instance, the German chemist Johann Rudolf Glauber worked with substances obtained from both the inanimate and animate worlds. From the mineral kingdom, he prepared crystals of sodium sulfate, in which each molecule, we now know, is loosely associated with ten molecules of water. Such a loose association is symbolized in the formula by a dot, and "Glauber's salt" (as it is still sometimes called) has the formula $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$. From the world of life, Glauber prepared such substances as acetone and benzene.

All these substances can be changed by gentle heating. There is a distinction, however. Glauber's salt will lose its loosely attached water molecules if heated, leaving behind the "anhydrous" sodium sulfate. If to this anhydrous substance water is added, however, crystals of Glauber's salt, identical with those originally heated, are re-formed.

On the other hand, if either acetone or benzene is heated, it will catch fire and not all the king's horses and all the king's men could help anyone restore the fiery vapors to the original acetone or benzene by any means as simple as cooling or adding water.

It was as though the chemicals obtained from the inanimate world shared the inertia and stability of the rocks from which they were obtained, while the chemicals obtained from living organisms shared the delicacy and fragility of life. Or we can put it in reverse and decide that rocks were dead and insensible because they were constructed of dead and insensible components, whereas organisms made up of unstable and fragile components were therefore themselves changeable, delicate, and sensitive.

By 1807, Berzelius put this distinction into words by speaking of substances obtained from living beings, or from the dead remnants of living beings, as "organic." Substances from the inanimate world were "inorganic." In accordance with the notion that there were two sets of natural laws,

one for life and one for non-life, chemistry was divided into the grand divisions of "organic chemistry" and "inorganic chemistry."

Since organic chemicals were produced only in organisms and never in the inanimate world, it seemed natural to suppose that they involved and, indeed, even required the spark of life (a "vital force") for their formation. If this were correct, then it followed that the chemist with his flame and his bellows and his acids could not manufacture an organic compound out of an inorganic one, for in which of his tools or agents lay the vital force that was required for the task?

The most important proponent of this viewpoint was Berzelius himself and it must therefore have come as a shock to him when his pupil, the German chemist Friedrich Wöhler, upset the applecart in 1828. Wöhler was an inorganic chemist, who was one of the first to prepare reasonably pure aluminum and beryllium, and who devised new methods for preparing nickel. At the time of his crucial discovery he was engaged in heating ammonium cyanate, a substance generally considered as thoroughly inorganic. Wöhler found, quite to his own amazement, that in the process of heating the properties of the ammonium cyanate changed completely. What he had instead was a substance with all the properties of urea, a thoroughly organic substance that was the chief dissolved material in mammalian urine.

An organic substance, then, had been formed out of an inorganic one, and by simple heating. No life force (except that in Wöhler himself, of course) had been required. The experiment was repeated over and over before Wöhler quite dared announce it, but once he did, it was confirmed, and, as frequently happens in science, other chemists rather promptly began to report the synthesis of other organic substances from inorganic sources.

Wöhler's discovery was not important so much from a chemical standpoint, for the chemical reaction he had reported was a rather trivial one in itself, as from a psychological standpoint. He had broken down what had seemed to be an absolute barrier between the animate and inanimate. The theory of the "vital force" was badly dented, and scientists were readier to believe that, in some fundamental

way, the same laws of nature applied to all objects, living or non-living.

With the certainty of difference torn away, various similarities between the living and non-living, already noted by scholars when such notice was unfashionable, now began to play a prominent part in scientific thinking. The chief of these involved the function of air.

A man, and all other land animals that could be studied by early scholars, had to breathe air in order to live, and the air breathed had to be reasonably fresh. In an enclosed place, animals grew lethargic and eventually died, as though in breathing they had used up some vital portion of the air (even though the air seemed unchanged to the naked eye).

This is reminiscent of the behavior of fire. A fire blazes up when a current of air is sent through it, and it flickers, lowers, and eventually dies when kept in an enclosed place. (To be sure, it is most often an organic substance such as wood or coal that is burning and perhaps it might be argued that life is still indirectly involved. However, the same is true if either hydrogen or sulfur, both completely and indisputably inorganic, burns.)

As long ago as 200 A.D., the Roman physician Galen had speculated that an understanding of the principles of combustion would lead to an understanding of those of respiration, and his prediction began to come true with the work of Lavoisier.

When Lavoisier demonstrated, in the late 1770s, that air consisted of nitrogen and oxygen, he showed plainly that it was the oxygen that supported both combustion *and* life. Furthermore, he went on to show that expired breath was not only comparatively poor in oxygen but also comparatively rich in carbon dioxide. More careful study, however, convinced Lavoisier that the carbon dioxide present in the expired breath did not account for all the oxygen decrease. There must be something else, besides carbon, with which the oxygen was combining.

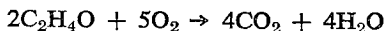
When, in 1781, the English chemist Henry Cavendish showed that when hydrogen burned in oxygen water was produced, the matter was suddenly clear to Lavoisier. Expired breath was certainly moist and the molecules in food certainly contained hydrogen. In the process of respiration,

oxygen must combine with hydrogen as well as carbon and must produce water as well as carbon dioxide. When wood burned, the same thing happened; oxygen was consumed and water and carbon dioxide were produced.

Doubtless it would have been daring to make too close a comparison between the animate and the inanimate at the time, but after Wöhler had synthesized urea and the boundary between the two worlds had been dissolved, the comparison grew unavoidable.

With the new belief that the laws of organic chemistry were similar to, or even identical with, those of inorganic chemistry, interest in organic chemistry increased explosively. Chemists began to analyze organic substances and to make use of the newly developed atomic theory to attempt to work out formulas for them, just as they were working out formulas for inorganic substances.

It began to appear quite clear that the organic substances in food, for instance, consisted largely of carbon, hydrogen, and oxygen, in the over-all proportion (roughly) of two atoms of carbon to four of hydrogen to one of oxygen. (There were also present minor quantities of nitrogen and still smaller quantities of sulfur, phosphorus, and so on, but these can be disregarded for the present.) We can imagine an equation such as the following as representing the type of reaction that goes on in the human body, with food being symbolized as C_2H_4O :



The constitution of wood is not too different from that of food generally (though perhaps $C_2H_4O_2$ might be a fairer formulation in its case) and wood, too, combines with oxygen to produce carbon dioxide and water. Clearly, then, from a chemical viewpoint both the body and the steam engine obtain their energy from the same basic source: from the union of carbon and hydrogen atoms with oxygen.

This, in itself, is a sobering thought to those who like to consider the body as something apart from and infinitely superior to a mechanical contrivance. Yet, to be sure, the similarity is not absolute. Burning wood produces heat and light, while the food "burning" in the body produces at most only a very gentle warmth and no light at all.

The question might arise, then, whether this much difference might not betoken much more basic differences. Since

the human body (and the living organism generally) produced energy in forms different from that of a flame, might it not do so much more efficiently? Could living tissue not rise superior to the limits placed on the inanimate steam engine by men such as Carnot? Must the body, too, obey the laws of thermodynamics that were being worked out in the nineteenth century, or could it, in its clever way, find a source of energy in food that the steam engine could never find in wood or coal?

To decide that, let's look at the nature of food more closely.

From early times, man could not help but notice that in preparing his food he was involved with several types of materials of widely different properties.

There was, for instance, the white, tasteless, easily powdered substance he obtained from such grains as wheat or rice, of which the chief component is now called "starch." (This is apparently derived from an Anglo-Saxon word meaning "strong," perhaps because it could be used to stiffen textiles, to the joy of generations of women and the discomfort of generations of men.)

There was also the sweet-tasting substance in fruit juice and in honey. From plant juices such as that of certain canes that are particularly rich in the substance, "sugar" itself could be prepared in solid form. This was first accomplished in India perhaps as early as 400 B.C. In ancient times some of it may have reached Rome, where the Latin name *saccharum* was originally derived, in all likelihood, from the original Hindu word. It was not until the Middle Ages that Europe really received sugar in any quantity and that by way of the returning Crusaders.

With the nineteenth century came the realization that there was more than one kind of sugar. In 1802 the French chemist Joseph Louis Proust (working in Spain) reported a substance obtained from grapes that certainly looked like sugar, being white, crystalline, sweet, and water-soluble; but yet was not ordinary sugar, for it was neither quite as sweet nor quite as water-soluble as ordinary sugar. He distinguished therefore between "cane sugar" and "grape sugar." Another sugar, which a decade before had been isolated from honey, and which was even sweeter than cane sugar, Proust called "fruit sugar."

It is now customary to give to the various sugars names that end in the suffix "-ose." Cane sugar is "sucrose" and fruit sugar "fructose" (merely Latinized versions of the common names). As for grape sugar, that eventually received the name "glucose" from the Greek word for "sweet," although it is the least sweet of the three. A sugar isolated from milk (and occurring nowhere else) is called "milk sugar" or "lactose" from the Latin word for "milk."—Still, when the word "sugar," unmodified, is used, it always refers to sucrose.

In 1812, as I said in the previous chapter, Kirchhoff made the rather amazing discovery that starch could be converted to grape sugar when it was heated with a bit of acid. Its molecule therefore can be viewed as being made up of a number of glucose molecules somehow bonded together. The act of boiling in acidified water broke those bonds, liberating free glucose.

By 1819 the French chemist M. H. Braconnot had boiled sawdust, linen, tree bark, straw, and a few other plant materials in acidified water and found that glucose was liberated in each case. The substance in these inedible products of the plant kingdom that, like starch, was built up of glucose units was nevertheless not starch, precisely because it was inedible. It required a separate name. Since the material turned out to be the stiffening substance surrounding and protecting plant cells, it was named "cellulose" by the French chemist Anselm Payen. (This name set the fashion, as a matter of fact, for the "-ose" endings.)

The French chemist Joseph Louis Gay-Lussac undertook to analyze the chemical composition of sugar and the related substances. He found that for each carbon atom present in the molecule there were present two hydrogen atoms and one oxygen atom. It seemed to him clear that, since a combination of two hydrogen atoms and one oxygen atom made up a water molecule, sugar must consist of a number of carbon atoms to each of which a water molecule was attached. Therefore he called the entire group of substances—starch, cellulose, the various sugars—"carbohydrates" ("watered carbon"). The molecules of these substances turned out to have a more complicated structure than that, but the name persists and will, undoubtedly, persist indefinitely.

As it turned out, the molecule of glucose contains six carbon atoms and so its formula can be written $C_6H_{12}O_6$,

and we can take this as a formula typical of the carbohydrates.

Another type of substance obtained from foodstuffs since earliest times were the fats and oils. Fats are softly solid while oils are liquid but in other respects the two are similar and are, indeed, lumped under the term "lipids" from a Greek word for "fat." The lipids are insoluble in water and have a distinctive greasy feel.

At about the time Braconnot was investigating the chemistry of carbohydrates, another French chemist, Michel Eugène Chevreul, was investigating the lipids. From his work there eventually came the realization that lipids, like carbohydrates, were composed of combinations of the three elements, carbon, hydrogen, and oxygen. The over-all difference lay in the fact that comparatively little oxygen was present. A typical lipid molecule, as we know it today, would have the formula $C_{57}H_{104}O_6$.

Still a third component of food, quite different from either carbohydrate or lipid, is to be found, in solution, in the white of egg (or "albumen" from a Latin word meaning "white"). Such substances form solutions which, like albumen, are thick and sticky. When heated gently, the substance coagulates and will not thereafter redissolve. Mild heat has changed its properties permanently.

A whole family of "albuminous substances," yielding viscous solutions and almost always coagulating on heating, had been isolated by the end of the eighteenth century. These included substances from blood and milk, and eventually from plant products as well.

What most set off this group from carbohydrates and lipids was the discovery that albuminous substances were not made up of carbon, hydrogen, and oxygen only, but that they contained nitrogen as well and, often, sulfur and even phosphorus. This was sufficient to indicate that the formula of albuminous substances was more complicated than those of the other classes of foodstuffs and, indeed, all attempts in the early nineteenth century to devise such a formula failed.

The Dutch chemist Gerardus Johannes Mulder did indeed try, in 1838, to devise a unit molecule, made up of a mere eighty-eight atoms, and to demonstrate that different albuminous substances could be built up of varying numbers of this, plus combinations of small sulfur- and nitrogen-con-

taining molecules. This device failed completely but nevertheless managed to introduce an important word into the vocabulary. At the suggestion of Berzelius (an inveterate word-coiner), Mulder called his unit molecule "protein" from a Greek word meaning "of first importance" because he considered the unit to be of first importance in the structure of albuminous substances. The name was eventually applied to the whole molecule and is now universally used in this sense.

By 1827 the English chemist William Prout (not to be confused with the contemporary French chemist, Proust) was already suggesting that the organic matter of foodstuffs, and of organisms for that matter, was made up essentially of the three classes of substances that now we would call carbohydrates, lipids, and proteins. This has been shown to be indeed the case and has been accepted ever since. These three classes do not, in actual fact, make up *all* of tissue. There are inorganic components, notably in the bones, and, of course, water. There are also organic components, such as certain substances called "nucleic acids," which are of key importance to life and which fall into none of the three classes.

Nevertheless, carbohydrates, lipids, and proteins make up almost all the organic components of tissue and we can concentrate on them.

The analysis of organic substances in the first half of the nineteenth century raised distinctions of an entirely new sort between organic and inorganic substances. For one thing, all the substances obtained from organisms that possessed the characteristic properties of organic substances contained carbon and hydrogen atoms in the molecule. Usually oxygen was also present; often nitrogen; sometimes sulfur and others.

On the other hand, while typical inorganic substances often contain hydrogen, oxygen, nitrogen, sulfur, and many other types of atoms in their molecules, they rarely contain carbon atoms. (An example of a compound, usually considered inorganic, that does contain carbon is limestone, or calcium carbonate.)

This explained the inflammability of organic compounds, for both carbon itself and, of course, hydrogen are inflammable. Any compound with a molecule largely constructed

of carbon and hydrogen is likewise inflammable. It is the chemical constitution that is important in this connection and not the mystical influence of some vital force.

A second distinction between organic and inorganic rests in the size of the molecule. Even the largest inorganic molecules are relatively small, consisting of not more than a dozen atoms or so. On the other hand, only the smallest organic molecules have so few constituent atoms. Glucose, a comparatively small organic molecule, contains 24 atoms, while a typical lipid molecule contains 170. Starch and protein molecules contain thousands and even millions of atoms.

Again, it is because of the large and rickety structure of the organic molecule that they decompose on heating, and change their properties, often, just on gentle warming; and again not through any mystical connection with life.

The proof of this came in the 1850s and thereafter, when chemists learned to synthesize not merely compounds identical with those isolated from tissue, but myriads of other large-molecular compounds built up of carbon, hydrogen, and other elements, compounds that did not occur in nature. These synthetic compounds, never a part of any living object, resembled in properties those that were. They were inflammable, delicately unstable, and in every way "organic."

Gradually, then, as the nineteenth century wore on, the distinction between organic and inorganic came to be oriented about pure chemical factors, and no longer hinged on life and non-life. Since the building up of the large molecules characteristic of organic compounds depended on the peculiar properties of the carbon atom (as was eventually found), the simplest distinction was to consider organic chemistry as dealing with carbon and all molecules containing carbon, while inorganic chemistry dealt with everything else.

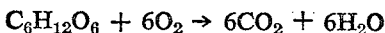
This division may seem unequal and so it is, but not in the sense that would appear on the surface. The ability of the carbon atom to join together in combinations of nearly limitless variety makes the total number of known compounds of this one element far greater than that of compounds of all other elements combined, and the disparity increases with each year.

With the chemical nature of food at least roughly blocked

out in this manner, it became possible to interpret properly quantitative measurements involving respiration.

Lavoisier, for instance, whose great service to chemistry consisted of insisting on accurate measurements, was the first to try to determine the exact quantities of oxygen consumed and carbon dioxide given off in respiration. He lacked the proper equipment for completely accurate results, but even if he had obtained accurate results he would not have been able to do much with them except to record them.

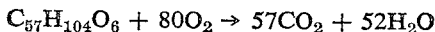
Now, however, with something learned about the molecules to be found in food, there was sense to be made out of the oxygen that disappeared and the carbon dioxide that appeared in its place. Suppose, for instance, that we consider glucose as a fuel; suppose we burn it in a flame. The equation representing such a combination of glucose and oxygen, properly balanced, would be:



From this equation (which could not have been written until after the structure of glucose had been worked out) you can see that for every mole of glucose consumed, six moles of oxygen are used up and six moles of carbon dioxide produced. For every mole of one gas that disappears, a mole of another gas appears. It so happens that a mole of any gas fills just about the same volume, so that what we are saying is that for every bit of volume of one gas that disappears that same volume of another gas appears. To be specific, for every liter of oxygen consumed, a liter of carbon dioxide is produced.

The ratio of carbon dioxide produced to oxygen consumed is, then, 1/1, or 1. Lavoisier and others were interested in this ratio in connection with respiration, so its value has come to be called the "respiratory quotient," a phrase that is usually abbreviated R.Q. We can say, then, that the R.Q. of glucose (and, indeed, of carbohydrates generally) is 1.

For lipids the matter is different. The typical lipid molecule has a formula $\text{C}_{57}\text{H}_{104}\text{O}_6$ and the equation representing its combination with oxygen, properly balanced, is:



For every 80 moles of oxygen consumed, 57 moles of carbon dioxide are produced. The R.Q. of lipids, then, is, in this case, $57/80$ or 0.713. The R.Q. of proteins is intermediate, between those of carbohydrates and lipids. In fact, the rounded values would be these:

R.Q. carbohydrates	= 1.0
R.Q. proteins	= 0.8
R.Q. lipids	= 0.7

Now how could these purely chemical deductions be related to life phenomena? The answer came in 1849, when two French chemists, Henri Victor Regnault and J. Reiset, devised a box into which animals could be placed and into which measured amounts of oxygen could be led. The carbon dioxide produced by those animals in the process of respiration could be collected, absorbed in appropriate chemicals, and weighed. In this way Regnault and Reiset were able to measure, by appropriate calculations, the volume of oxygen consumed by the respirations of the experimental animal, and the volume of carbon dioxide produced. They could, in other words, measure the R.Q. of the animal, and the animals they experimented with ran the gamut from earthworms to dogs.

Regnault and Reiset found the R.Q. of living organisms to lie within the limits 0.7 to 1.0. Furthermore, the value shifted with the nature of the food. When the diet consisted primarily of carbohydrate, the R.Q. approached 1.0; when it consisted primarily of lipid, it approached 0.7.

A couple of decades later the German chemist Max von Pettenkofer and his coworker, the physiologist Karl von Voit, designed an apparatus large enough to hold a man. The R.Q. in man possessed values like those found for the lower animals. It, too, varied between 0.7 and 1.0 according to diet. On a normal mixed diet, or when the subject was fasting and consuming his own reserves of food in some normal balance of carbohydrate and fat, the value was just about 0.8.

By mid-century, then, it was quite clear that the over-all course of combustion of organic substance was the same whether it went on outside living tissue or within it. At least the proportions of carbon dioxide produced and oxygen consumed were just those predicted from chemical reactions

following the rules that had been worked out through a study of inanimate nature. The body seemed to be subject to the law of conservation of matter at least. It could not create carbon out of nothing, or, for that matter, do away with a jot of it.

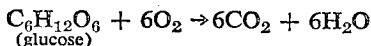
But that still left the law of conservation of energy; a law that was just coming into acceptance at the time Regnault and Reiset began their first experiments. Would that law, too, hold for life as for non-life?

Chapter 14

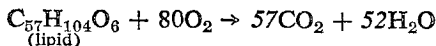
THE IDLING SPEED

In considering the energy balance in living organisms, the first step is to determine the amount of chemical energy liberated by the combustion of food under conditions where life is *not* a factor. Different food components can be burned in a bomb calorimeter and the heat of reaction can be measured. In this case, however, it would be inconvenient to try to determine the molar heat of reaction, since some of the more important components of food have such large molecules that the molar heat of reaction would be astronomical in size.

If we take only the simpler foodstuffs, we can set up the following equation, including the molar heat of reaction (see Chapter 8):



$$\Delta H = -738 \text{ kilocalories}$$



$$\Delta H = -8044 \text{ kilocalories}$$

As you see, we must already deal with values in the thousands of kilocalories. For starch and protein, with their

giant molecules, matters are far worse. A mole of starch or protein could easily have a mass of several hundred pounds and the molar heat of reaction involved in burning such a mass would be in the millions and even billions of kilocalories. Besides, until well into the twentieth century, there were no accurate means of determining the molecular weight of these giant molecules, so that even if experimenters had a lust for large numbers, it would have done them little good.

The heat of reaction can, however, be measured quite conveniently on a per-gram basis. The combustion of carbohydrates in a bomb calorimeter yields a heat of reaction of 4.1 kilocalories per gram. This value is quite low as compared with some of the combustibles I mentioned in Chapter 8. To refresh your mind, the heat of reaction per gram is given for a few of those substances in Table 4.

Table 4

HEAT OF COMBUSTION PER GRAM

Substance	Formula	Heat per Gram
Hydrogen	H_2	34.2 kcal.
Methane	CH_4	13.2 kcal.
Coal	C	7.8 kcal.
Ethyl alcohol	C_2H_5O	7.1 kcal.
Glucose	$C_6H_{12}O_6$	4.1 kcal.

As I explained in Chapter 8, hydrogen yields much more heat per gram, when burned, than carbon does, and a hydrocarbon such as methane would naturally yield an intermediate quantity of heat. The fact that alcohol delivers even less heat per gram than carbon itself, despite the presence of hydrogen atoms in its molecule, I explained as due to the presence of an oxygen atom as well. The alcohol molecule was already "partially burned" and less free energy remained for delivery as heat when burning was completed. Following this line of argument, glucose (which we may take as a typical carbohydrate), with its formula $C_6H_{12}O_6$, is more "burned" than is alcohol. One quarter of the atoms of the glucose molecule are oxygen, while only one ninth of the atoms of the alcohol molecule are. Consequently, it is not

surprising that on a per-gram basis glucose delivers less heat when burned than alcohol does.

The case is different for lipid. The lipid molecule, with only about 1/28 of its atoms oxygen, is far less "burned" than carbohydrates or even than alcohol. The heat of reaction per gram developed by the combustion of lipid should therefore be higher than for alcohol, but lower than that for methane, the molecule of which contains no oxygen at all. Experiments bear this out, for the combustion of lipid produces 9.45 kilocalories per gram. Protein, with molecules that contain less oxygen than do those of carbohydrates and more than do those of lipids, is intermediate and produces 5.65 kilocalories per gram when burned in a calorimeter.

So much can be obtained by making a bonfire out of food. What about the results if the food is used as fuel not for a fire but for a living organism? An organism is harder to handle than a flame but it can be done. What is needed first is a calorimeter large enough to hold an organism. The organism eats and excretes wastes; inhales oxygen and exhales carbon dioxide. From the details of eating and breathing, if measured accurately, the amount of food substance combined with oxygen within the tissues of the organism could, ideally, be calculated. This amount of combination would produce a known quantity of heat if it were conducted in a calorimeter. If the heat produced by the organism is measured, it would then prove to be more than or less than or just equal to the amount that would have been produced by ordinary burning. Which of these alternatives is true is of vital importance if we are to decide whether the laws of thermodynamics hold for living organisms.

This sort of energy experiment was first accurately carried through by the German physiologist Max Rubner in the 1880s. Rubner went about it most methodically. In the first place, he carefully analyzed the feces produced by the experimental animal. Even when the animal was fed thoroughly digestible lipid and carbohydrate, some small quantity of it always passed through the alimentary tract unabsorbed, and in any fair test this had to be subtracted from any food the animal ate, since unabsorbed food is not available for energy. It turned out that very little carbohydrate survived alimentary passage, but that rather more lipid did. Allowing for failure in absorption, the energy content of ingested carbo-

hydrate and lipid could be reduced to just about 4 kilocalories per gram and 9 kilocalories per gram respectively.

Protein introduced a new complication. Whereas the final products of the combustion of carbohydrate and lipid are carbon dioxide and water, whether that combustion is carried on in the calorimeter or in the body, the same is not true of protein. In the calorimeter the carbon and hydrogen atoms in the protein molecule are converted to carbon dioxide and water, but the nitrogen atoms must also be taken into account. They are liberated as a nitrogen oxide. (There are also atoms of sulfur and other elements in many protein molecules, but they are present in such small quantities that they can be ignored in this connection.)

In the mammalian body, however, the nitrogen content of proteins is converted not to nitrogen oxide molecules but to urea (CH_4ON_2), a soluble solid that is discharged in the urine. The molecule of urea will produce some energy on combustion, so when it remains unconsumed and is excreted as is, it represents a loss of some of the total energy of combustion of the protein molecule. Subtracting the energy of urea from that of protein makes it clear that the energy available to mammalian tissue from protein is just about equal to that from carbohydrate—4 kilocalories per gram.

The fact that urea is produced turns out to be very useful, however, as I shall now explain. The total amount of foodstuff consumed and "burned" by tissue in a given time can be determined by measuring the carbon dioxide produced in that interval. However, the amount of energy liberated in this manner depends upon the nature of the foodstuff consumed. Carbon dioxide arising from the combustion of lipid represents more energy than the same amount of carbon dioxide arising from the combustion of carbohydrate.

What Rubner found it necessary to do was to measure both the oxygen consumed and the carbon dioxide produced. That gave him the R.Q. and from that he could tell the proportion in which carbohydrate and lipid were being consumed in the body. The higher the R.Q. the greater the preponderance of carbohydrate utilization in the body and the less the energy being produced per liter of oxygen consumed. The lower the R.Q. the greater the preponderance of lipid utilization in the body and the greater the energy being produced per liter of oxygen consumed.

This, however, works accurately only if carbohydrate and

lipid are the only foodstuffs being consumed in the tissues and if protein is *not* being consumed. Nevertheless, this is an unfair assumption because protein *is* being consumed, usually, as we now know, to an extent of about 10 per cent of that of the combined carbohydrate and lipid. Unless the exact amount of protein involved is known, the R.Q. figure cannot be relied upon to give good values for the energy available to the organism.

Thus, suppose the R.Q. happens to be 0.8. If lipid and carbohydrate are the only foodstuffs being consumed, then they are being consumed in the ratio of 2 parts lipid to 1 part carbohydrate and the energy production, on the basis of bomb calorimetry, can be expected to be about 7 kilocalories per gram.

Suppose, however, that the actual mixture being consumed were 2 parts lipid, 1 part carbohydrate, and 1 part protein. The R.Q. would still be 0.8 (since that is the R.Q. of pure protein, the addition of pure protein in any quantity would not affect the R.Q. of a two-to-one lipid/carbohydrate mixture). However, the energy production of this new mixture would go down to about 6 kilocalories per gram, because protein is relatively poorer in energy than the carbohydrate-lipid mixture.

However, only protein produces urea. From analysis of the urine (another innovation by Rubner), the amount of protein being consumed in the tissues can be calculated. By feeding his animals weighed quantities of food, by analyzing feces and urine, by measuring oxygen consumption and carbon dioxide production, Rubner was able to determine exactly how much of each type of foodstuff was being consumed by an animal over a period of time. From bomb calorimeter experiments he knew how much energy should be produced from these consumed foodstuffs.

By measuring the amount of heat produced by the organism, he (and later experimenters, too) found that the expected happened. The organism produced as much heat out of the consumed foodstuffs as would have been produced in a bomb calorimeter, no more and no less.

By the turn of the century it was quite accepted that living things had no magic source of energy; they could get no more energy out of a system than there was energy to be obtained. In short, the first law of thermodynamics held for the lizard as firmly as for the rock.

The fact that the calories of food consumed and of energy produced must balance in a living organism does not mean that the value of both cannot vary together by large amounts. Calorimeters have been devised that are large enough to contain human beings who are engaged in a variety of activities: standing, sitting, typing. Indirect methods have been devised to measure energy expenditure while running, climbing stairs, bicycling, horseback riding, and so on.

It is no surprise that, as activity grows more strenuous, energy production goes up and so does the quantity of fat and lipid that must be consumed, either through larger meals or through a greater draw upon the body reserve with consequent loss of weight.

In fact, in any question of losing or gaining weight, the law of conservation of energy is implicitly accepted. To lose weight, it is necessary to cut calorie intake below expenditure, either by eating less food (dieting) or by expending more energy (exercise) or both. To gain weight requires the opposite.

There are both upper and lower limits to the rate of energy expenditure. By running upstairs three at a time, or chopping wood furiously or playing tennis with all one's might, energy can be expended at the rate of 10 kilocalories per minute over a short period of time. We all know this sort of thing cannot be continued for long.

On the other hand, by suspending all voluntary activity, the rate of energy expenditure can be cut down to a fifteenth that quantity, but no lower. No matter how we relax and make ourselves comfortable, a certain amount of energy must be expended to keep the heart beating, the lungs pumping, the liver and kidneys working, and so on. This minimal energy expenditure can be compared to that of an automobile engine idling at a speed just great enough to prevent stalling. (Of course, there is an obvious reminder here that we must be careful how we go about making parallels. The automobile engine can be stopped, then started again. The human "engine" can be stopped easily enough, but cannot be started again.)

The lowest rate of energy expenditure takes place, naturally enough, when a person is asleep, but it is not very practical to attempt to measure the energy expenditure of a sleeping person. For one thing, there is considerable variation, since

there is such a thing as restless sleep, and it is difficult or impossible to tell a subject: "Sleep quietly."

The next best thing is to remain quietly awake, and even though the minimum rate of energy expenditure under wakeful conditions is 10 per cent higher than under the most favorable sleeping conditions, wakefulness is perforce accepted.

The chemical changes that take place in the body and, among other things, produce the energy that is being measured, is referred to as "metabolism." This term was introduced in 1839 by a German naturalist, Theodor Schwann, and is taken from a Greek word meaning, literally, "throw into a different position" and, therefore, implying "change." The minimal energy expenditure I have just been referring to, the "idling speed" so to speak, must be balanced by a minimal, or basal, rate of metabolism. The minimal energy expenditure is therefore referred to as "basal metabolic rate," a phrase that is usually abbreviated BMR.

The conditions under which BMR is determined require first that the subject be lying down. (The mere act of sitting up requires additional energy because the large muscles of the trunk must keep up a continual interplay of contraction to maintain balance.) The subject must be fasting, not having eaten for at least twelve hours prior to the test, to make sure no additional energy is being utilized to maintain the digestive processes. The room must be comfortably warm to make sure that no extra energy is being utilized for the production of abnormal amounts of body heat (if the room is too cold) or perspiration (if the room is too warm). The subject must be mentally relaxed as well to prevent the undue consumption of energy involved in the muscular tension that accompanies mental tension. (The best way to achieve relaxation is to allow the subject to lie quietly for half an hour in silence and peace.)

The BMR can then be taken most accurately by measuring oxygen consumption, carbon dioxide output, and the urea content of urine over a period of time. More simply, one can ignore the urine and suppose the protein consumption to be insufficient to affect the results materially. More simply still, one can measure oxygen consumption only and assume that the R.Q. is going to be near the normal 0.82 for a fasting subject.

A rough figure for the BMR for the average human being turns out to be about 43 kilocalories per hour or about 1000 kilocalories per day. This varies widely with weight, as is to be expected, since the more tissue there is the more energy must be expended to keep the total mass alive. Thus, whereas a man weighing 115 kilograms (250 pounds) has a BMR of 68 kilocalories per hour, and one of 70 kilograms (150 pounds) has one of the aforesaid 43 kilocalories per hour, a boy weighing 30 kilograms (65 pounds) will expend only 20 kilocalories per hour.

You would expect less variation if you calculated the BMR per kilogram body weight, and if that is done, it turns out that the 115-kilogram man expends energy at the rate of about 0.59 kilocalories per kilogram body weight per hour, while the 70-kilogram man expends energy at the rate of about 0.61 kilocalories per kilogram body weight per hour. The 30-kilogram boy, however, expends energy at the rate of 0.67 kilocalories per kilogram body weight per hour. This is not surprising really, for the boy is growing and the chemical processes involved in laying down new tissue are energy-consuming.

(Rubner, in 1883, found that the BMR could be made even more constant if it were related not to the body weight but to the surface area of the body.) *Since BMR is body weight*

The determination of BMR of other mammals offers no surprises. Mammals larger than man have a total BMR that is larger and those smaller have a total BMR that is smaller. On a weight basis, however, the BMR goes down as the size of the animal goes up, just as it does for human youngsters as compared with human adults.

For instance, while a man's BMR is 1000 kilocalories per day, that of a mouse is only 3 kilocalories per day, and that of an elephant is about 50,000 kilocalories per day. However, one kilogram of mouse tissue consumes 158 kilocalories each day, while one kilogram of man consumes 15 kilocalories and one kilogram of elephant consumes only 10 kilocalories.

There is always a temptation to remember that large animals, on the whole, live longer than small animals do, and one cannot help but wonder if longevity has something to do with metabolic rate. Can it be that small animals, living at a faster pace, wear out sooner?

The trouble with this theory is that there are too many exceptions, too many cases where the smaller animal outlives the larger. The most notorious of these exceptions is man himself.

Man lives longer than any of the great apes, including the larger-than-human gorilla. For that matter, man lives longer than the colossal elephant. There is no reliable record of any elephant ever having reached the age of seventy, even when under the best care; whereas, as we all know, half the human beings in the advanced areas of the world live to be seventy and centenarians are not unheard of. Even whales, whose longevity is difficult to study, probably do not live as long as men do, since even the largest whales become mature enough to bear children within three years of birth, and longevity span correlates with the age at which biological maturity is attained. Compare the whale's case with the maturation age of thirteen for man.

In fact, the only animals of any kind known to outlive men are some of the larger tortoises, where ages approaching two hundred seem to be possible. But then tortoises are notorious for their slow rate of living. For fast living over a number of years (in more than one sense perhaps) nothing can approach man.

The basal metabolic rate might perhaps have been nothing more than a statistic if it had not been found to vary sharply under interesting circumstances. To explain this, I shall have to detour by way of the word "gland."

The word is derived from the Latin *glans*, meaning "acorn," and was originally applied to any small scrap of tissue within the body, such as the "lymph glands." These are small, ovoid bodies lying along the course of some of the vessels carrying lymph (the fluid that leaks out of the tiniest blood vessels and bathes the cells of the body). The acorn is a natural metaphor to use for anything small, after all, since people have always been impressed with the smallness of the original acorn in comparison with the largeness of the final oak.

The practice spread, and larger organs with a tapering, acorn-like shape, such as the pancreas, were included under the name. In 1642 a German anatomist, Johann Georg Wirsung, discovered a duct leading from the pancreas into

the upper end of the small intestine. This, in turn, led to the discovery that the pancreas secreted a fluid or juice into the intestine.

Such was the impression made by this discovery that people began to think of glands as those organs whose main function was to produce and discharge juices. For that reason, any organ that did this, however large and however un-acorn-like its shape, received the name of "gland." For instance, the bile-producing liver (which weighs up to four pounds in the human being) is called a gland. On the other hand, scraps of tissues that do not produce juices, however small and acorn-like they might be, tend to lose the name. The lymph glands, for which the name was originally invented, do not produce fluids and are now usually called "lymph nodes" from the Greek word for "knot," since the swellings do, indeed, resemble knots in the long strings that are the lymphatic vessels. Nevertheless, the old name dies hard and when the nodes swell in response to some infection (which it is one of their functions to fight), mothers and doctors alike speak of "enlarged glands."

The juice-secreting glands first recognized were, naturally, those which, like the pancreas and liver, had obvious ducts through which those juices reached their destination. Other glands of this type are the sweat glands, the mammary glands, the small glands in the lining of the stomach and intestine, and so on.

With the nineteenth century, however, it was realized that some small organs produced fluids that traveled through no ducts at all but were discharged directly into the blood stream. There was some dispute as to whether such organs ought to be called glands or not, but the liberal view won out. The German anatomist Johann Friedrich Meckel divided glands into two groups, those with a duct and those without, a division that persists to this day. The "ductless glands" are also called "endocrine glands" ("separate within"), since their secretions are separated from them within the body, instead of being led to the skin or the alimentary canal and taken out of the body itself.

One of the ductless glands half surrounds the trachea (or windpipe) just below the larynx (or voice box) (see Fig. 21). The larynx itself is made of cartilage and has a notch in front, which you can feel with your finger if you place it to your throat. This notch resembles the notches in the shields

used by Greek soldiers in Homeric days. (The notch was there in order that the soldier might get a quick view of the enemy while keeping his body protected.) The Greek word for such a shield was "thyreos" and so the cartilage composing the voice box was called the "thyroid" ("shieldlike") cartilage. In 1646 the English physician Thomas Wharton transferred the name from the cartilage to the little fleshy organ near it, so that this is now known as the "thyroid gland."

The thyroid gland is sometimes enlarged, swelling out the neck with what is called a "goiter" (from the Latin word for "throat"). This was viewed with considerable complacency in ancient and medieval times, and in those areas of Europe where goiter was common the mild swelling of the neck was accepted as an addition to female beauty. By 1800, though, physicians were noting that goiter was sometimes accompanied by severe symptoms of disease, of which the most noticeable, if not the most serious, were protruding eyes.

An undersized thyroid was found to be even more alarming. Certain mountain regions in Europe seemed to breed instances of dwarfish, feeble-minded individuals called "cretins." (This is a dialectical form of the French word for "Christian," the sufferers being so called in contempt or, perhaps, pity, as we might speak of a "poor soul.") This, too, was found to be connected with the thyroid. In fact, toward the end of the nineteenth century it was found that the removal of the thyroid in animals would induce cretin-like symptoms that could be relieved by injecting extracts of mashed-up thyroid.

Exactly what it was about the thyroid that affected the body so remained a complete mystery until 1896, when the German chemist Eugen Baumann discovered that the thyroid contained the rare element, iodine. The total content was so small that the discovery was underrated. Physicians and biochemists, at the time, had not yet realized that it was possible for elements to occur in the body in minute quantities ("trace elements") and yet be essential to life.

Nevertheless, in 1905 an American physician, David Marine, settling down in the Midwest after his training in the East, noticed that goiter was widespread among men (and animals, too) in Cleveland, whereas he had noticed little of this in the East. He wondered if this might be due to the fact that there was less iodine in the soil, hence in plant food, hence in livestock, hence in man, near Cleve-

land than on the east coast, which was subjected to the iodine-containing ocean spray. If thyroid required iodine for

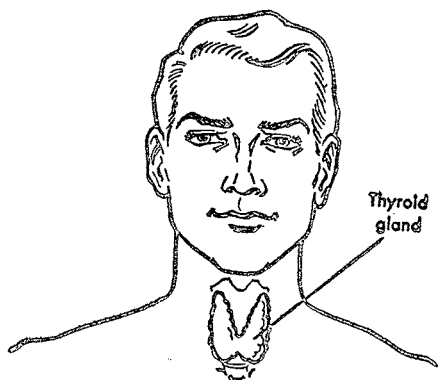


Figure 21. *The Thyroid Gland*

proper functioning, perhaps it grew overlarge in compensation for iodine lack yet still slowed down in activity. Or, if it gained necessary iodine while overlarge, its activity might rise too high.

Marine found that by feeding animals on an iodine-free diet he could induce goiter, which he could then relieve by adding iodine to the diet.

He began to campaign for the addition of iodine-containing compounds to the city's drinking water, and after a long period of resistance (very similar to the resistance now offered against the similar proposal to add fluorine-containing compounds to drinking water) succeeded. In this way, and by the later use of salt to which small quantities of iodine-containing compounds have been added ("iodized salt"), the dangers and discomforts of goiter have greatly diminished in advanced areas of the world. Those dangers are not yet entirely gone, however. It has been estimated that even now, a half century after the discovery of iodine therapy for goiter, there are 200,000,000 people in the world who suffer from that condition. Of these, 30,000,000 are in Latin America and there is even a notable incidence in certain parts of the United States, particularly in the states of Washington and Oregon.

Meanwhile, just as the chemical attack was revealing the

importance of iodine to thyroid functioning, clinical research discovered something of allied importance. In 1895 the German physician A. Magnus-Levy applied the then new technique of BMR determination to people suffering from various diseases. For the most part, he obtained normal values. Significant variations did occur, however, in those patients suffering from thyroid abnormalities. Where the thyroid was overactive ("hyperthyroid") the BMR was markedly higher than normal; where it was underactive ("hypothyroid") the BMR was markedly lower than normal.

From this arose the concept that the thyroid gland controlled the "idling speed" of the human body; that it was a sort of energy thermostat. (Exactly how it exerts this control is not understood even today.) The use of BMR was now plain; it was no longer a mere physiologic curiosity but became an important clinical tool for the detection of thyroid abnormalities. So common did the use of this tool become that "hyperthyroid" and "hypothyroid" became part of the general vocabulary.

A hyperthyroid individual with a BMR more than 10 per cent above normal tends to be overactive, tense, nervous, and overemotional. A hypothyroid individual with a low BMR would be dull and listless. By the use of thyroid extract or by anti-thyroid substances, it has become possible to raise or lower thyroid activity at will.

After a half-century reign, however, the BMR test, conducted as I have described it earlier in the chapter, is becoming a thing of the past.

The reason for this dates back to 1914, when the American biochemist Edward Calvin Kendall isolated from thyroid extract a relatively small molecule that he named "thyroxine." This proved to be the active principle of the extract. The active principles of the endocrine glands had, in 1903, been named "hormones" by the English physiologist Ernest Henry Starling. (The term comes from a Greek word meaning "to arouse" because of the marked stimulating effect of the endocrine secretions on various organs of the body.) Thyroxine was therefore the thyroid hormone.

The molecule of thyroxine is unusual in that it contains, as part of its structure, four iodine atoms. In 1952 the British biochemists J. Gross and R. V. Pitt-Rivers discovered another thyroid product, "tri-iodothyronine," which differed

from thyroxine only in possessing three rather than four iodine atoms in its molecule. It, too, had thyroid hormone activity. Either way, though, thyroxine or tri-iodothyronine, it is the iodine atoms that are characteristic, and they occur virtually nowhere else in the body.

As the thyroid functions, it releases its hormone into the blood stream. There it attaches itself to some of the protein molecules normally found in the plasma (the liquid portion of the blood). If the proper protein fraction is separated, it could conceivably be analyzed for its iodine content. This would indicate the amount of hormone being produced by the thyroid (since the hormone is the only source of the iodine), which would, in turn, indicate the level of activity of the thyroid, and this, it follows, would give the BMR. In this way the determination could be carried on without fasting, without lying still for an hour, without breathing through a hose, without elaborate rooms and equipment—with no more trouble than the withdrawal of some blood.

The catch is that analyzing the iodine in blood is far easier said than done. Something like $1/20,000,000$ of a gram has to be detected with reasonable accuracy. By the middle 1950s this was achieved.

In order to detect so small a quantity, use is made of the phenomenon of catalysis. Once the "protein-bound iodine" (usually abbreviated PBI) has been broken down and the iodine collected in solution, possibilities for a direct determination are slim. However, to the solution are added known quantities of chemicals that will react under the catalytic influence of the iodine. The rate at which the reaction proceeds will vary with very minute changes in iodine concentration so that measurement of the rate offers a most delicate test for the iodine.

The course of the reaction can be followed easily by color changes that can be measured by appropriate instruments. The more rapid the color change, the faster the reaction, the greater the concentration of iodine, the higher the level of activity of the thyroid gland, the more the value of the BMR.

Indirect, but very convenient!

Chapter 15

THE CATALYSTS OF LIFE

It is all very well to equate combustion and respiration and to look at the human being as a bonfire, fanned by oxygen, and producing carbon dioxide. Yet there is no hiding the very important differences that exist between a living being and a bonfire.

For one thing a bonfire moves in only one direction; the wood or other combustible that feeds it is converted to carbon dioxide and water, leaving ash as a residue. A bonfire will not re-create its own fuel. Living tissue, however, which will subsist on and obtain its energy from carbohydrate, fat, and protein by reactions analogous to combustion, will at the same time store fresh carbohydrate and fat and build up new protein to replace the old as I shall describe in some detail later in the book. If the organism is growing, it will even build these complicated molecules at a rate greater than that with which it consumes them, so that a ten-year-old boy, for instance, contains more of each of these substances than a seven-year-old does, despite having busily consumed them for three years.

The building of the complicated molecular components of the tissues involves (as we shall see in more detail later) an increase in free energy and therefore a decrease in entropy; a reversal, in short, of the route of spontaneous reaction predicted by the laws of thermodynamics.

As far back as the end of Chapter 5, I suggested that the distinction between life and non-life might involve the ability of the former to bring about local decreases in entropy. At the time, I spoke only of the decreases in entropy involving mechanical energy (that is, in "making an effort"), and cited the characteristic ability of so many living things to move upward against gravity. Now it appears that life can also bring a decrease in entropy in terms of chemical energy; it can

bring about the formation of large complex molecules from simpler ones although the direction of spontaneous reaction is from large to small.

That strengthens the connection between life and entropy decrease, but it *still* does not suffice for a good distinction. I said in Chapter 5 that the heat of the sun in evaporating the oceans and setting air in motion brought about a local decrease in entropy, and that forces within the earth in raising mountains did the same thing and that in neither case was there any form of life (by any useful definition) brought into play.

To be sure, these non-life decreases in entropy involve only mechanical energy, but such decreases involving chemical energy can also be cited. Irradiation of solutions of simple molecules with ultraviolet light will result in the formation of more complex molecules, and there is every reason to think that this took place on a very large scale during the youth of the earth, when life was nonexistent in its oceans and the sun's ultraviolet blazed down upon it for millions of years. Again, no form of life would be involved. In fact, that non-life decrease of entropy in connection with chemical energy may have been the very means by which life itself first came into being.

So the distinction we are seeking still eludes us.

However, there is no need to be disheartened. We can search for the distinction between life and non-life in another obvious difference between a bonfire and living tissue.

A bonfire is hot as blazes (literally) and a human being is not. The human being is warm to be sure, but the normal body temperature is 37°C. , which is simply not to be compared with the 700°C. , and more, to be found in flames. (What's more, cold-blooded animals consume oxygen and produce carbon dioxide down to temperatures of nearly 0°C.)

This at once makes it necessary to consider reactions analogous to combustion being initiated at low temperatures. There is no possibility of supposing that the body's fuel is any easier to ignite than the ordinary fuel of a bonfire. The various foodstuffs, kept at room temperature outside the body (under sterile conditions, to prevent bacterial contamination), will remain unchanged indefinitely. There will be no perceptible combination with oxygen, any more than would

be the case for wood or coal. Within the body, however, the combination proceeds with merry ease at temperatures never surpassing the mildly warm.

The clear solution to this paradox, by the nineteenth century at any rate, was the existence within living tissue of catalysts, working after the fashion I described in Chapter 12. To be sure, the catalysts that brought about combustion in the laboratory, such as platinum and palladium, do not occur in living tissue, but that only means that catalysts of other types must occur. The question arises, though, whether tissue components analogous to catalysts are really catalysts in the non-life sense, obeying the laws of thermodynamics as blindly as powdered platinum metal would; or whether they are somehow bound up with life and "vital force" and rise superior to thermodynamics.

On this question, the chemists of the nineteenth century clashed, and, thanks to certain reactions known to man since prehistoric times, the weight of the evidence ran heavily (for a time) in favor of those who held out for a "vital force."

Fruit juices or steeped grain, on standing, will generally undergo changes of various types. The changes may not always be desirable but on occasion, as prehistoric man discovered, drinking the changed liquid can bring about a pleasant feeling of gaiety and warmth. Eventually men found that it was not necessary to expose the juice and hope for the best, but that transfer of a small quantity from an already pleasantly changed material to a fresh batch would hasten the change, and just the proper change, too.

In the same way, dough out of which bread was to be baked might occasionally undergo a chemical change that would bring about formation of carbon dioxide. The trapped bubbles of gas would fluff up the dough and produce a light, tender loaf quite different from the hard, flat dough that would ordinarily form. Again the transfer of some of the fluffed-up dough to a fresh batch would hasten the proper change in the new material as well.

The material that brings about the changes is called "yeast," a word that traces back to a Sanskrit word that means "boiling." This makes sense since the formation of gas bubbles in the manufacture of both wine and bread

resembles a kind of boiling without heat. An alternate name (used in the Bible) is "leaven," from a Latin word meaning "to raise," since the use of yeast raises the dough.

The process whereby dough was raised or fruit juice was turned to wine and grain mash to beer was, in early modern times, named "fermentation" from a Latin word meaning "to boil."

The action of yeast would, to a nineteenth century chemist, resemble the action of catalysts very strongly. What's more, it would differ from the ordinary catalysts of the laboratory in that it had the properties of an organic substance. Furthermore it catalyzed reactions involving just the sort of substances (the starch of grain or the sugar of fruit juices) that were to be found in tissue. If there were catalysts of any sort to be found in living organisms, surely it was reasonable to suppose they bore a resemblance to yeast rather than to platinum.

But yeast resembled something more than merely a catalyst; in some ways it resembled a living thing. It was not that it merely remained unconsumed, as an ordinary catalyst would; it actually increased and multiplied in the course of the reaction. In his first epistle to the Corinthians, the Apostle Paul takes note of this when he makes use of what probably was an old proverb even at the time: "Know ye not that a little leaven leaveneth the whole lump?"

It does; and when the whole lump is leavened, a small bit of it will suffice to leaven another whole lump and so on.

Despite this, however, it was not easy to jump to the notion that yeast was alive. The concept of living things invisible to the eye (angels, demons, fairies, and so on) was common enough among the ordinary folk of early times but the notion of ordinary non-mystical organisms, invisible only in the sense that they were too small to see, was harder to accept.

But with the seventeenth century came the invention of the microscope, and in 1675 the greatest of the early microscopists, the Dutch merchant Anton van Leeuwenhoek, saw and reported upon a variety of "animalcules." These were one-celled organisms (now called "protozoa" from Greek words meaning "first animals") a fiftieth of an inch long and less, yet as vital and alive as you and I. The discovery electrified the scholarly world of the time and the foreign

tradesman was elected a fellow of London's exclusive group of gentleman scientists, the Royal Society.

In 1680, Van Leeuwenhoek (who looked at everything through the tiny, perfect lenses he ground himself with infinite patience and love) looked at yeast under the microscope and found it to consist of tiny spheroids. The thought of life might have occurred to anyone now that microorganisms were known, but the spheroids of yeast were quiescent and showed none of the obvious signs of life shown by the actively swimming protozoa.

In 1837, however, the French physicist Charles Cagniard de la Tour, observing yeast under the microscope (much improved by then over Van Leeuwenhoek's simple lenses), caught the little spheroids in the act of budding (see Fig. 22). This was reproduction and reproduction meant life.

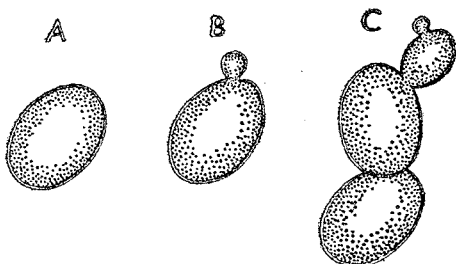


Figure 22. *The Progressive Budding of Yeast Cells from A to C*

Yeast was established as a one-celled living thing, more nearly allied to the plant than the animal kingdom and therefore vegetative and quiescent.

In the 1850s and 1860s the French chemist Louis Pasteur, in a series of classic investigations, showed that fermentation is generally the result of the action of living microorganisms. Undesirable products, such as sour wine, were produced by strains of yeast other than those responsible for good wine. Again, the corruption and putrefaction of meat and of the dead remains, generally, of once living tissue, he finally proved (though others had preceded him with less conclusive demonstrations) to be the result of the activity of living microorganisms.

In short, where it might have been thought that examples of organic catalysts had been discovered, it turned out that

examples of life had been discovered. The vitalists were heartened and they recovered from the blow given them by Wöhler's synthesis of an organic substance in the laboratory.

To be sure, they could not take up their old position of the absolute distinction between organic and inorganic on the basis of the "vital force," but there was a good second line of defense. They were forced to admit that the chemist could synthesize organic compounds from inorganic ones without the intervention of life, *but* only by certain routes. They had to use high temperatures and pressures, or strong chemicals, or catalysts not present in living tissue. Only life itself, they maintained, could form those organic compounds, and particularly the very complicated ones, at ordinary temperature and pressure and with only the mildest of chemical conditions.

But another line of attack on the catalysts of life offered itself. Next to the gaseous interchange of oxygen and carbon dioxide in respiration, the chemical processes associated with life that could be most easily investigated were those of digestion.

The reason for this is that the body, essentially, is doughnut-shaped. The alimentary canal is the hole in the body-doughnut, with openings to the outer world at the mouth and anus. For something to be located in the hole of a doughnut is not the same as its being located in the doughnut itself and, analogously, a process like digestion, which takes place in the alimentary canal, is taking place outside the body.

The practical significance of this is that samples of the stomach contents can be drawn out easily enough through the throat without seriously affecting the subject.

As long ago as 1752 Réaumur, of the Réaumur temperature scale, investigated the digestive processes of a falcon. He wanted to find out if the stomach's action was one of mechanically grinding food into a paste or if more subtle changes were involved. He therefore managed to get the falcon to swallow small metal tubes, containing scraps of meat. The metal tube protected the meat from any grinding action. The tubes were capped at either end with wire net that would permit any stomach juices to enter.

When the falcon regurgitated the tubes (such birds ordinarily bolt food whole and regurgitate indigestible residue)

Réaumur found the meat dissolved, without having been ground, and the tube filled with clear fluid. Nor was there any odor of putrefaction. He decided that some chemical action other than putrefaction had taken place, a kind of fermentation.

Other investigators in following decades reported the stomach contents to be acid and, in 1824, William Prout proved that hydrochloric acid was the particular acid present. Acid was known, by then, to be a good catalyst for such reactions as the breakup of starch and proteins to smaller molecules, and it was reasonable to think that this might be the answer to digestion: acid catalysis.

However, as investigations proceeded, substances were isolated from gastric juice, which, when dissolved, could bring about the dissolution of meat and yet were not acid. The German naturalist Theodor Schwann, in 1835, obtained a powder that was not acid and that was strongly active in breaking down meat. He called it "pepsin" from a Greek word meaning "digestion."

Here was something unusual. Pepsin was at once a catalyst and something that was thoroughly organic, or seemed to be. It became fashionable to call such an organic catalyst a "ferment" because digestion and fermentation, both allied with life, were being looked upon as much the same sort of process.

Meanwhile, two French chemists, Anselm Payen and J. F. Persoz, had done something similar in connection with an extract from malt (steeped barley). In 1833 they had obtained a ferment that could bring about the conversion of starch to a sugar called "maltose." They called this new ferment "diastase." (Payen had good luck with his names. His term "cellulose" had made fashionable the ending "-ose" for carbohydrates. Now his term "diastase" eventually made fashionable the ending "-ase" for ferments, although not until after a considerable period during which digestive ferments were given names ending with "-in," after the example of pepsin.)

The isolation of various ferments continued. The German chemist, Justus von Liebig, along with Wöhler, studied an extract of bitter almonds that had previously been reported as bringing about the breakdown of a plant substance called "amygdalin." By 1837 they had decided that the extract contained a ferment, which they named "emulsin." Then the

French chemist Auguste Pierre Dubrunfaut managed to get an extract from yeast that would bring about the breakdown of sucrose to the simpler sugars, glucose and fructose. This he called "invertase."

What's more, these ferments did not break the laws of thermodynamics, it was found, any more than did ordinary catalysts. Ferments might alter the rate of reaction but not the position of the equilibrium point (see Chapter 9). In this respect, too, the body seemed to have no special powers as compared with the inanimate universe.

This sort of thing rocked the vitalist position but did not destroy it. The isolated ferments were organic, yes, and they could catalyze certain reactions, to be sure, but they catalyzed only those reactions (so it seemed) that resembled digestive breakdown. And digestive breakdown took place outside the body tissues and could not therefore be considered a true life process.

What the isolated ferments could not do, the vitalists maintained, was ever to catalyze reactions that went on only within the tissues themselves. For instance, the isolated ferments could not convert sugar solutions to alcohol. That could only be brought about by intact living yeast cells, and the energy relationships within the cell still remained in question.

To stress this point, they categorized ferments into those involved with life processes and those not. The former were "organized ferments." The latter, like pepsin or invertase, were merely "unorganized ferments."

In 1876 the German physiologist Wilhelm Kühne went a step further. He suggested that the term "ferment" be reserved for life processes only. The unorganized ferments, acting in the absence of life, he called "enzymes" from Greek words meaning "in yeast" because they could be obtained from yeast, among other sources.

Thus the vitalist position maintained itself, by clever retreats, throughout the nineteenth century. And this position implied always that there was this vital force characteristic of life that acted under laws different from those obtained through a study of inanimate nature.

But then, in 1897, a German chemist, Eduard Buchner, wished to prepare a whole yeast extract for study and to do so with as little damage as possible. He decided to grind the yeast cells thoroughly with sand and then to filter the

mash. He obtained in this way a clear yeast juice containing no living cells.

By the vitalist notions of the day, this yeast juice was dead and therefore completely incapable of fermenting sugar. Therefore Buchner felt quite safe in preserving the juice against bacterial attack by adding a strong solution of sucrose—a common preservative. To his utter amazement, however, the added sucrose proceeded to ferment. Bubbles of carbon dioxide appeared and so did alcohol.

This process of fermentation, supposedly so characteristic of life, had been brought about by thoroughly dead chemicals, and the entire vitalist position collapsed, never to recover. Organized ferments behaved just as unorganized ferments did from the energy standpoint. The laws of thermodynamics applied outside the cells, inside the cells, everywhere. There was no further retreat open.

Ferments might be produced by life, but they were not themselves living, and their action was independent of life. From that time on, the term "enzyme" was applied to all ferments, inside and outside of cells, organized or unorganized. In fact, the term "ferment" dropped out of the scientific vocabulary altogether.

Merely to give a name—enzyme—to the organic catalysts associated with life solved no problems. A name is not an explanation.

What agitated the minds of biochemists at the turn of the century was the question of the exact chemical nature of the enzymes. A catalyzed reaction usually contained only very small quantities of catalyst (in the experience of those working with inorganic catalysts), so that it seemed quite possible that enzyme extracts might catalyze reactions powerfully and yet contain too little enzyme for actual study. This proved, indeed, to be the case. Furthermore, any extract from living tissue contained so many and such varied substances, of sometimes very complex nature, that it would task the patience of Job and the wisdom of Solomon to decide which, if any, of them was the enzyme.

About the only way of attacking the problem, in the beginning, was indirect. The presence of the reaction meant the presence of the enzyme and the absence of the reaction meant the absence of the enzyme. Any condition that would stop the reaction without any apparent effect on the reacting

chemicals might therefore be explained as being a condition that had rendered the enzyme inactive.

For instance, gentle heating at a temperature well below the boiling point of water usually stopped an enzyme-catalyzed reaction after a few minutes. This was not due to any effect on the reactants, for if the reactants were heated, then cooled, and if an enzyme solution were then added, the reaction would start at once. However, if the enzyme solution were heated and then cooled, and if it were added to the reactants, there would be no reaction. The effect of heating, then, clearly involved the enzyme molecule, which was, apparently, quite fragile, at least in its resistance to heat.

One type of molecule was known to display exactly this sort of sensitivity to heat, and that was the protein molecule. Could it be, then, that enzymes were proteins? The evidence began to pile up. Other environmental conditions known to break down the complex protein molecule (such as increased acidity, or mechanical shaking of solution) also put an end to enzyme-catalyzed reactions. Chemicals known to combine tightly with proteins put an end to enzyme-catalyzed reactions. Other enzymes, known to catalyze the breakdown of proteins, often put an end to enzyme-catalyzed reactions. It began to be taken for granted, in the early twentieth century, on the basis of such indirect evidence, that enzymes were proteins.

This view received a setback about 1920, on the basis of apparently direct evidence advanced by the German chemist Richard Willstätter. He purified enzyme solutions to an unprecedented pitch, freeing them as far as possible of any substances other than the enzyme itself. He ended with a clear solution that had a great deal of catalytic activity (and therefore possessed the enzyme) but which, apparently, contained no protein. He used all the chemical tests employed in those days for the detection of proteins, and all ended negative. No protein.

It might have occurred to him that the protein tests of that day were of only limited sensitivity and that the enzyme itself might be present in such small quantity that it would not be detected by such tests even if it were protein. This attitude, however (although subsequently shown to be correct), was not taken by Willstätter. His conclusion was that, whatever the enzyme molecule was, it was not protein. He was a renowned chemist, a Nobel prize winner, and his

viewpoint had great influence. His position was accepted in the decade that followed.

Meanwhile, though, an American biochemist, James Batchellor Sumner, was working with ground-up jack beans. From them he could extract a solution that would catalyze the breakdown of urea to ammonia and carbon dioxide. The enzyme involved was, very naturally, given the name "urease" in consequence. From the solution of urease, Sumner was able to obtain small crystals which, when separated and dissolved in fresh water, turned out (somewhat to Sumner's surprise) to have strong urease action.

Sumner did everything he could think of to separate the activity from the crystals but could not succeed. Anything that destroyed or damaged the crystals ended the activity, and Sumner eventually decided he actually had the enzyme itself in crystalline form. Since the crystals checked out as protein by all tests, his conclusion was that urease, at least, was a protein.

Sumner's discovery made little headway at first against Willstätter's reputation, but in 1930 another American biochemist, John Howard Northrop, announced that he and his associates had crystallized several enzymes, including pepsin. Other crystallizations followed in rapid succession, and in every case the crystals were protein in nature.

The Willstätter theory broke against the weight of the evidence and in the 1930s it was firmly and finally established that enzymes were protein in nature. It is now possible to feel perfectly secure in defining enzymes, in all briefness, as "catalytic proteins" or "protein catalysts."

Chapter 16

THE DETAIL OF PROTEIN

In the light of the previous chapter, we can now once more try to distinguish between life and non-life; between the lizard and the rock. Suppose we try this:

A living organism is characterized by the ability to effect a temporary and local decrease in entropy by means of enzyme-catalyzed chemical reactions.

That the ability is temporary is a recognition of the fact that the lizard will not remain alive for more than a few years even if all its needs are met to the full. In fact, only a very few animals (man and some turtles) seem ever to live as long as a century, and only a very few plants (certain trees) can ever live as long as a millennium, whereas a non-living object such as a rock can maintain itself essentially unchanged for several billions of years certainly and, presumably, for an indefinite period beyond.

That the ability is local is a recognition of the fact that the decrease in entropy is not an over-all phenomenon in defiance of the second law, but that it comes about at the expense of a greater increase in entropy elsewhere in the system of which the living organism is a part.

That the ability is effected through enzyme-catalyzed chemical reactions is the crucial part of the definition. It is that which distinguishes a living organism from systems that bring about local decreases in entropy through the use of sources of energy such as the radiation of the sun or the internal heat of the earth. It also distinguishes a living organism from man-made machinery, which may effect local decreases in entropy through use of heat engines, chemical cells, or other methods not involving enzymes.

To a certain extent, this distinction is arbitrary. Why insist on the use of enzymes? Human beings are now constructing computers and automated factories that seem to be aping more and more of the traditional attributes of life and even of human life. Even such abstractions as thought, judgment, learning would seem to be involved. Granted that no man-made object even today approaches the full complexity of life, might it not someday? And when it does shall we refuse to grant it the accolade of "life" because it makes no use of enzymes? Or what if men isolate certain enzymes (or even synthesize them) and make use of them in certain key steps in the workings of such machines? Would the machines then be alive?

Attacking the problem from another direction, and one which, thanks to the events of the last few years, is no longer fantastic, what if we come across objects on other planets that possess all the attributes we are accustomed to associate

with life; including, perhaps, even intelligence? Suppose also that the basis of their existence is not the enzyme, or even proteins generally, or even, for that matter, chemical reactions? Are they not to be considered alive?

In considering these questions, I find myself a little regretful that we do not accept the vitalist position today. If the vitalists were right and if there existed two sets of laws of nature, one for living systems and one for non-living systems, it would be possible to evolve a completely unambiguous distinction. A living organism would be one that follows the first set of laws; a non-living object one that follows the second set.

It is precisely because the vitalists are considered to be wrong—and that it seems as certain as anything can be in science that one set of laws governs both living and non-living systems—that a distinction between the two is difficult to make. To make the distinction too broad would always run the danger of defining a star or a solar system as alive. To make it too narrow always runs the danger of making life too purely anthropomorphic.

Some people who have considered this problem have cut the Gordian knot by deciding that distinction between life and non-life is meaningless. To attempt such a distinction, they feel, is merely to bow to our own prejudices in the matter; to our own preconceived notion that there is such a thing as life on the one hand and non-life on the other. In actuality, there may only be organizations of matter at different levels of complexity, with "life" an arbitrary name given to certain levels and "non-life" to certain other levels, and with no clear boundary between.

I am in sympathy with this last view, but I must admit that it leaves us rather helpless. Arbitrary distinctions need not necessarily reflect the Truth with a capital T, provided only they can be useful to us in our thinking and that they are recognized as arbitrary. Thus, the distinction between organic and inorganic chemistry is purely arbitrary and is kept because it is Useful and not because it is True.

Therefore, recognizing its arbitrariness and its possible incompleteness in the face of man's continuing inventiveness and the results of his forthcoming forays into space, I should still like to advance the statement in italics at the beginning of this chapter as an adequate and useful distinction between life and non-life; brief, simple, and clear, and holding for all

cases likely to be encountered by you and me at this moment in man's history.

Having finally found (to my satisfaction and, I hope, yours) the distinction between life and non-life which, at the beginning of the book, we set out to find, we have reached one of our goals. It would be frustrating, however, to be satisfied with this accomplishment alone, for it leaves unanswered some very fascinating questions.

After all, chemical reactions and catalysts are found in plenty in systems that are undeniably non-living. The reactions in non-living systems and in living organisms are basically similar and follow the same set of natural laws. Why, then, should a mere change in catalyst effect all the tremendous difference that strikes us when we compare life and non-life?

There is no denying that a bee is a more intricate mechanism than a jet-plane and a mouse a more wonderful construction than a mountain.

If it is the existence of the enzyme that makes the difference, then let's look at the enzyme closely, and try to find out why.

In any comparison between enzymes and the ordinary catalysts characteristic of non-living systems, various points of difference force themselves on our attention. Enzymes are much more fragile and easily inactivated; they are also much more demanding in the conditions under which they will work. Temperature and acidity must be just right, certain ions must be present in small quantities, while others must be rigidly excluded, and so on.

Again, enzymes are highly specific in the reactions they will catalyze. Whereas ordinary catalysts such as platinum and acids will hasten any of a wide variety of reactions, particular enzymes are quite limited in the nature of the reactions they will catalyze. It is not uncommon for an enzyme to catalyze one and only one reaction, while possessing no effect at all on other reactions, however similar to the "one and only one" they may be.

Matching these differences in behavior (which may be of degree rather than of kind, but which are startling enough even so) is a difference in structure. Enzymes are proteins and other catalysts are not. It seems natural to suppose that

the unusual properties of enzymes are the outgrowth of the nature of the protein molecule.

Let's start then with that.

Through the nineteenth century, what was most clear about the protein molecule was that it was large.

Thus, in 1860, the Scottish chemist Thomas Graham found that if he separated a solution of some substance in water from a sample of pure water, by means of a thin strip of parchment, the molecules of the dissolved substance could often find their way through the separating membrane. They could then be easily detected, by appropriate chemical tests, in what had previously been pure water.

This did not happen for every substance in solution. Sodium chloride and other soluble inorganic substances passed through the parchment. So did such soluble organic substances as the various sugars. However, a solution of protein remained on its own side of the parchment. The protein molecule did not pass through. Graham gave names to the two classes of materials, those that passed through the parchment and those that did not. Since those that passed through happened to consist of substances that easily formed crystals upon evaporation of the solution, he called them "crystalloids." On the other hand, one of the protein solutions he tested was glue (which in Greek is *colla*), so he called the substances that did not pass through the parchment "colloids."

The simplest explanation of all this (and one that, with refinements, still holds) is that there are submicroscopic holes in parchment that will allow passage of small molecules but not of large ones. As a result, parchment and other membranes of this sort—especially including the membranes that enclose living cells—are called "semi-permeable membranes."

This makes possible one of the most common methods of purifying solutions of proteins. The protein solution is placed in a tied-off bag composed of a semi-permeable membrane (usually sausage casing, these days) and is placed in running water. All small-molecular-weight substances pass through the membrane and are washed away, leaving only the large protein molecules behind. This process is known as "dialysis" (see Fig. 23).

The fact that proteins will not pass through a semi-

permeable membrane sets a lower limit on the size of the protein molecule. The fact that they remain in solution at

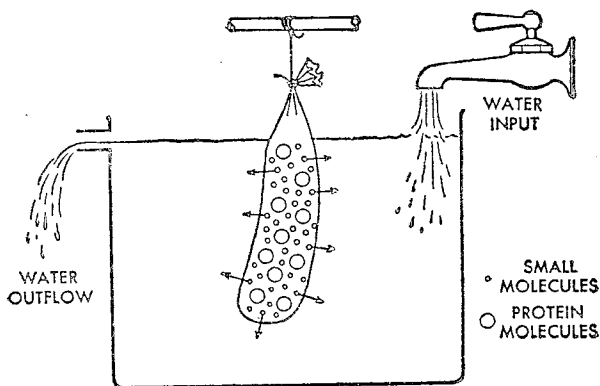


Figure 23. *Dialysis through a Semi-permeable Membrane*

all sets an upper limit. Let me explain why the latter is true.

Every particle suspended in water is acted upon by two opposing forces. First, there is the force of gravity acting to pull the particles downward (supposing them to be denser than water, which they usually are) and, second, there is the random bombardment of water molecules, which tends to keep them mixed homogeneously through the solution. (If the particles carry electric charges, there are electrostatic forces that may make them clump together or stay apart, but we can ignore that for the present.)

For any particle of visible size, the gravitational force is the stronger, so that the particle sinks downward. However, as the particle size decreases, the molecular bombardment effect becomes more important and the rate of settling is correspondingly slowed. We can see that for ourselves if we stir up a handful of soil in a container of water. Any pebbles present settle out at once when the stirring is done. After that, somewhat smaller particles settle out, then still smaller particles, then fine particles. In the end, haziness may hang on as some particles persist in remaining suspended, being so small that the molecular bombardment effect reduces their

settling out ("sedimentation rate") to small values or even altogether to zero.

Particles barely small enough to remain in suspension indefinitely are too small to be seen under ordinary microscopes. However, they are large enough to scatter light, an effect discovered by the British physicist John Tyndall in the 1870s and called the "Tyndall effect" in consequence. If you view such a suspension at right angles to the course of a light beam being sent through it, you will see the beam marked out as a line of haze, the haze representing the small fraction of the light scattered toward your eye.

If the eye is replaced by a specially designed microscope (an "ultramicroscope") the particles become visible as bright dots.

For smaller particles, the light-scattering effect becomes weaker, until at the point where the particles are small enough to pass through a semi-permeable membrane the effect is too weak to see and the solution is "optically clear."

Protein solutions show the Tyndall effect. This combined with the fact that protein molecules are large enough to find a semi-permeable membrane impassable, and small enough to remain in suspension indefinitely, shows that those molecules are of "colloidal dimensions."

Particles with colloidal dimensions have diameters as large as ~~1000 millimicrons~~ at one extreme and as small as 1 millimicron at the other (a millimicron, remember, being a billionth of a meter). Even the very smallest colloidal particle would have to be made up of about a thousand atoms, while the largest would be made up of as many as a trillion atoms. Now the various atoms making up protein molecules have weights that average out, on the atomic scale, to 7. We can say, then, that the molecular weight extremes that would allow protein molecules to remain in the colloidal range must be 7000 and 7,000,000,000,000.

A range that wide requires a good deal of narrowing down before we can be satisfied, but it is sufficient to show that protein molecules are far larger than are molecules of such substances as sodium chloride (molecular weight $58\frac{1}{2}$) or glucose (molecular weight 180).

During the last quarter of the nineteenth century scientists began to tackle the problem of getting precise values for

the molecular weight of giant-sized molecules such as those of proteins. One way was to make use of dialysis. If a protein solution is on one side of a semi-permeable membrane and pure water is on the other, then water molecules can travel freely through the membrane in either direction. On the solution side, however, the large protein molecules, which cannot pass through the membrane pores themselves, block some of the pores, so to speak, through which water molecules would otherwise pass.

The result is that water molecules travel from the water side to the solution side more rapidly than they travel from the solution side to the water side. There is thus a net flow from the water side to the solution side. The water level rises on the solution side and falls on the water side. The weight of the water making up the difference in level exerts a pressure that is counter to the flow of water and eventually stops the flow altogether.

This flow from water to solution is an example of "osmosis" (from a Greek word meaning "to push") and the pressure that builds up is "osmotic pressure" (see Fig. 24). The size of the osmotic pressure depends on the number of large molecules blocking the pores; the greater the number of molecules, the higher the pressure.

Suppose, though, that you begin with a definite weight of protein. The larger the molecule of that particular protein, the fewer molecules into which that weight can be divided. Hence, for a given weight of protein, the lower the osmotic pressure, the higher the molecular weight.

The first to attempt to measure the osmotic pressures produced by various colloidal solutions was a German botanist, Wilhelm Pfeffer, in 1877. By 1885 the Dutch physical chemist Jacobus Hendricus van't Hoff had worked out the exact relationships between osmotic pressure and molecular weight, and a period of refinement of method set in.

By 1917 molecular weight determinations through osmotic pressure measurements had become quite reliable. In that year the Danish chemist Sören Peter Lauritz Sørensen reported that egg albumin (the original "albuminous material") had a molecular weight of 34,000. Shortly thereafter, the molecular weight of hemoglobin (the chief protein of red blood corpuscles) was reported as 67,000.

Unfortunately the osmotic pressure drops to such low values as molecular weight moves up into the range of a

hundred thousand and more that an accurate determination becomes impossible. For many important proteins, molecular

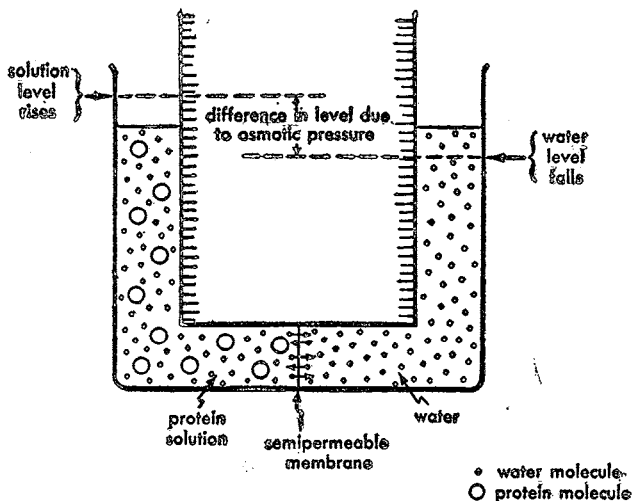


Figure 24. *Apparatus for Determining Osmotic Pressure and the Molecular Weights of Protein Molecules*

weight determinations could not be made. Nevertheless, by 1925 it was quite clear that the molecular weight of proteins was rarely less than 10,000 and that many important proteins had molecular weights under 100,000.

Another line of attack was in the direction not of diffusion through a membrane but of sedimentation. As I said earlier, the smaller the particle, the slower the sedimentation rate. Unfortunately protein molecules are all small enough to have a sedimentation rate of zero, so that the larger molecules cannot be distinguished from the smaller in this way—at least not under ordinary conditions.

In order to make the sedimentation rate become useful, one must do something that, in effect, strengthens the gravitational field so that the protein molecules *do* start settling out, each at a rate dictated by its size. To tamper with gravitation itself is, as yet, beyond man's powers, but it is possible to make use of another phenomenon to achieve the same result. If a container of liquid is whirled about

rapidly, the contents have a tendency to move away from the center of revolution and toward the portion of the container farthest from that center. This "centrifugal effect" is, in many ways, quite analogous to gravity and, like gravity, produces a sedimentation. Unlike gravity, the centrifugal effect can be easily increased by merely spinning the container more rapidly.

In 1925 the Swedish chemist Theodor Svedberg devised an "ultracentrifuge" that could be made to whirl so rapidly as to produce a centrifugal field in which protein molecules would sediment out of solution. (At present, in fact, ultracentrifuges capable of developing fields equivalent to 500,000 times that of ordinary gravity can be constructed.)

The sedimentation rate can be followed as the ultracentrifuge spins by means of ingenious optical devices. When more than one protein is present in solution, the sedimentation rate of each can be distinguished. (The rate of sedimentation depends on the shape of the molecule as well as its weight, for an elongated molecule settles out more quickly than a spherical one of the same weight. The molecular shape must also be determined by independent methods in order that sedimentation rate give reasonably accurate information as to molecular weight.)

By means of the ultracentrifuge, the range of experimentally determined molecular weight was moved upward. The protein fibrinogen (present in blood and responsible for the ability of blood to clot) proved to have a molecular weight of 330,000. Molecules of hemocyanin (a bluish protein molecule in the blood of sea creatures such as clams) have molecular weights ranging from 4,000,000 up, while the proteins making up viruses have molecular weights in excess of 10,000,000.

At the very moment, though, we are interested in enzymes, and these are not to be found among the very largest protein molecules. The molecular weight range of enzymes can be placed between 10,000 and 500,000. Very few enzymes will be found outside that range.

The large size of the enzyme molecule, as compared with that of ordinary catalysts, is not, however, in itself an explanation for the unusual properties of enzymes. Many other types of organic molecules have giant-sized molecules also.

Starch, cellulose, rubber, the many synthetic plastics, all possess giant molecules but none of these have enzymatic properties.

One important difference between proteins (including enzymes, of course) and other giant molecules is the relative ease with which most protein molecules are disrupted. It is easy to say that giant molecules are naturally rickety and can easily be made to collapse by the vibrations brought on by increasing temperature. But why does this affect proteins and not other giant molecules? Size alone does not necessarily involve fragility. Houses can be built out of the giant molecules of wood and can withstand an extraordinary amount of buffeting without deterioration.

Let's get deeper than mere size, then.

Two structures of equal size may differ in stability, depending on the nature of the units out of which they are built. A six-foot-high structure built out of wooden blocks is stability itself compared to an equally high and complex structure built out of playing cards. Can we compare the units out of which a protein molecule is built with the glucose out of which starch and cellulose are built and see what the difference might be?

It proved less easy to determine the structural unit of protein than that of other giant molecules, though an early start was made by the French chemist M. H. Braconnot. Braconnot had already broken down a variety of carbohydrates by heating them with acid, after the fashion of Kirchhoff, and in 1820 he turned to the protein, gelatin.

This, too, he found could be broken down by acid and out of it he obtained sweet-tasting crystals that he naturally thought to be a variety of sugar. He called them "sugar of gelatin." However, he soon found it possible to produce what was unmistakably ammonia (NH_3) from the crystals, which meant they contained nitrogen atoms and could not represent an ordinary sugar. The substance was eventually named "glycine," the stem coming from a Greek word meaning "sweet" and the "-ine" being the conventional suffix used by chemists for nitrogen-containing substances.

This did not solve the problem of the structural unit of proteins, however. Whereas most giant molecules seemed to have a single structural unit (discover it and the problem is solved), proteins did not. Other units, similar to glycine,

were discovered. Braconnot himself located one he called "leucine," and through the nineteenth century and even into the twentieth, new ones were added to the list.

The exact number of such structural units is not easily stated, since some of them occur only in a few proteins and it is hard to decide whether a particular unit, occurring only in an obscure protein, perhaps, ought to be included in the general list. One compromise is to consider the nineteen different compounds that are to be found among the breakdown products of almost every protein as the major units, so to speak. The names of these, in the order of discovery, are included in Table 5.

Table 5

THE MAJOR UNITS OF THE PROTEIN MOLECULE

Name	Year Discovered	Name	Year Discovered
Glycine	1820	Histidine	1896
Leucine	1820	Cystine	1899
Tyrosine	1849	Cysteine	1901
Serine	1865	Valine	1901
Glutamic Acid	1866	Proline	1901
Aspartic Acid	1868	Tryptophan	1901
Phenylalanine	1881	Isoleucine	1903
Alanine	1888	Methionine	1922
Lysine	1889	Threonine	1935
Arginine	1895		

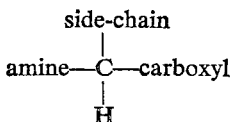
unlike

This is by no means as heterogeneous a list as might appear from the names alone. All these compounds have much in common. For instance, each one contains an atom combination called a "carboxyl group" which has acid properties. In addition, each one contains an atom combination called an "amine group." Because of this, Berzelius, as long ago as 1848, suggested that such compounds be called "amino acids" and the suggestion was adopted. The structural units of the protein molecule are, therefore, the various amino acids.

In each case the amine group and the carboxyl group are separated by a single carbon atom, so that the molecule of the various amino acids obtained from proteins can be looked on as having the following "skeleton":



In the case of glycine, the simplest (as well as the first discovered) of the amino acids, there are two hydrogen atoms attached to that central carbon atom and that is all the molecule. In the case of all the other amino acids, the central carbon atom is attached to one hydrogen atom and to one more or less complex carbon-containing group of atoms called a "side-chain." A more elaborate version of the amino acid skeleton can therefore be written as follows:

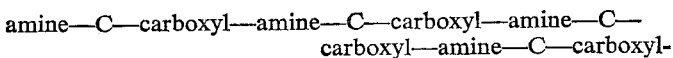


The amino acids, you see, are all similar in general (the skeleton) but are different in detail, for each possesses a different side-chain.

By the turn of the century, with a dozen amino acids isolated from proteins, the time was ripe to try to reverse the process and to form a protein out of amino acids. This problem was tackled by the German chemist Emil Fischer. Using the techniques of organic chemistry as developed over the previous half century, he painstakingly treated amino mixtures under such conditions as would encourage combination.

By 1907 he had managed to build up a molecule made up of eighteen amino acid units, consisting of fifteen glycines and three leucines. (Such relatively small strings of amino acids are called "peptides" from the Greek word for "digestion" because they are produced in the process of digestion.)

In doing this, Fischer laid to rest completely any possibility that the amino acids might not be the building blocks of proteins, for his eighteen-unit product had all the properties that might be expected of a very simple protein molecule. He was further able to show that, in forming a chain, the carboxyl group of one amino acid combined with the amine group of another, so that the skeleton of the peptides he formed was something like this:



The central carbon atom of the various bound-together amino acids carried each its own side-chain (or a hydrogen atom in lieu of one in the case of glycine) so that the peptide "backbone" of the protein molecule can be viewed as sprouting a series of varying side-chains of types and in an order depending on the nature and order of the amino acids in the chain.

The link between the carboxyl group and the amino group, first demonstrated by Fischer, is called a "peptide link" and accumulating evidence ever since his day has made it as certain as anything can be in science that naturally occurring proteins are built up of amino acids joined by way of peptide links.

Since Fischer's time, new methods for synthesizing peptide chains have been developed. A method developed in 1947 by the Israeli biochemist E. Katchalski, which uses somewhat modified amino acid molecules to begin with, results in the formation of peptides containing up to two hundred amino acid units.

The nature of the building blocks suffices, you see, to put protein into quite another class than that of almost any other kind of giant molecule. Starch molecules and cellulose are built up, essentially, of a single unit, that of glucose. The giant molecule of "inulin" (a starchlike substance produced by the Jerusalem artichoke) is made up of fructose units only. Rubber is made up of units of a hydrocarbon called "isoprene" and that only. "Hyaluronic acid," a slippery substance that lubricates our joints and is found between the cells generally, is made up of two different modified sugar units in alternation. Synthetic plastics are all built up of a single unit, or of two units in alternation.

One type of starch molecule may differ, then, from another in the number of glucose units it possesses; one type of inulin may differ from another in the number of fructose units it possesses, and so on. This allows for considerable differences in property between one giant molecule and its neighbor of the same type, far more than is possible in the realm of the small molecules—yet this room for difference

pales into utter insignificance in comparison with what is potentially possible for the protein molecule.

The protein molecule is built up of nineteen different amino acids (give or take one or two) and these may occur in any relative quantities in a particular molecule and in any possible order along the peptide chain. Each different arrangement of amino acids results in a different distribution of the various side-chains and therefore produces molecules of different properties in each case. The difference is not merely that one molecule may consist of a chain of eighteen amino acids and the next a chain of nineteen. Two molecules, each consisting of either eighteen or each of nineteen molecules, may still be different if the order of the particular amino acid molecules present differs.

How many varieties are thus made possible? Suppose you were building a protein molecule out of one each of the nineteen different amino acids. You could begin with any one of the nineteen, so you have nineteen possibilities right there. For each of these nineteen possibilities, you could have any of the eighteen remaining amino acids second in line, so that the total number of different combinations is 19×18 as far as the first two units are concerned. For each of these different combinations, you can have any of the seventeen remaining amino acids in third place, so the number of combinations is now $19 \times 18 \times 17$.

By continuing this chain of reasoning it turns out that the total number of different ways of combining nineteen different amino acids into a long chain is $19 \times 18 \times 17 \times 16 \times 15 \times 14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, a series referred to as "factorial 19" and symbolized as $19!$ Factorials mount rapidly as the number of integers in the series increase and quickly reach unimaginably high values, so that the exclamation point in the symbol would certainly seem to be justified. As an example, you may calculate the value of $19!$ or you may be willing to take my word for it that the value is over a hundred quadrillion—121,646,600,408,832,000 to be exact.

Well then, the possible molecules, each with different properties, that can be built up out of a mere nineteen amino acids number more than a hundred quadrillion.

But actual protein molecules contain far more than nineteen amino acids. The molecule of hemoglobin, with a mo-

lecular weight of 67,000, is made up of 539 amino acids. Not all of these are different, to be sure (a fact that cuts down the number of different arrangements that are possible), but even so, the number of possible arrangements of the amino acids in a molecule of hemoglobin comes to something like 4×10^{619} , or a 4 followed by 619 zeros.

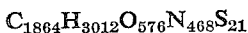
The significance of such a vast number is completely inconceivable. It merely points up the fact that the number of different proteins built up out of nineteen amino acids, each present in any quantity, is—like the number of different books that can be built up out of the twenty-six letters of the alphabet, or the number of different musical compositions that can be built up out of chord combinations involving eighty-eight notes—for all practical purposes, infinite.

The significance of such a number from the standpoint of energy considerations is, on the other hand, clear-cut and very interesting. It is relatively easy to put amino acids together blindly as Fischer or Katchalski did. That is equivalent to taking a handful of playing cards and simply stacking them without looking at them.

To be sure, the amount of order in a blindly stacked deck is higher than in a scattered deck, so that entropy is decreased in the stacking. (Remember the relationship between entropy and order described in Chapter 6.) The formation of a peptide from amino acids therefore involves a decrease in entropy.

The formation, however, of a particular protein such as hemoglobin out of all the virtually infinite numbers of amino acid combinations possible is the equivalent of stacking a large deck of cards in one particular arrangement and no other. The increase in order and hence the decrease in entropy is enormous. The feat of the body in forming protein is amazing not because it simply combines amino acids into a peptide chain, but because it combines them into *particular* peptide chains.

It is for this reason that there is little value in trying to find out the formula of a protein, if all one does is count the atoms. The result of such an endeavor, if carried through, would be something like this:



This is actually the formula, atom by atom, of a protein called "betalactoglobulin," which is found in milk. However, since such an atom count doesn't tell us anything about the kind of amino acids present, it tells us almost nothing of value. A protein molecule must be analyzed in terms of its amino acid content and not of its atomic distribution.

The straightforward method of analyzing a protein molecule for its amino acid content is to break down the protein molecule to an amino acid mixture, by cooking it gently with acid, for instance, and then analyzing the mixture for the quantity of each amino acid that it contains. Unfortunately the mixture is so complex and several of the amino acids are so similar in properties, as to make clear-cut separations of the mixture into individual amino acids almost impossible. Prior to 1940 the amino acid analyses of proteins remained incomplete and rather inaccurate, though broad outlines could be made out.

In 1944, however, two English biochemists, A. J. P. Martin and R. L. M. Synge, devised the technique of "paper chromatography." In this technique, a drop of the amino acid mixture is placed at one end of a piece of very porous paper ("filter paper") and is allowed to dry. The amino acid molecules remain firmly bound to a thin and invisible film of water on the paper. Now an organic liquid such as butyl alcohol (a near relative of the familiar ethyl alcohol) is allowed to creep up the paper, by capillary action. (You see capillary action whenever a blotter sops up a drop of ink.)

As the butyl alcohol passes the dried mixture of amino acids, each amino acid is subjected to two tendencies: one, to remain in the water layer and be motionless; two, to move into the butyl alcohol and move along with it. Each amino acid makes some sort of compromise. It moves along, but less rapidly than does the butyl alcohol. Furthermore, each amino acid makes a slightly different compromise (for each, thanks to its differing molecular make-up, is soluble to a different extent in water and in butyl alcohol). Each therefore moves at a different rate, and after a while each amino acid is to be found on a different spot on the paper. In this way the individuals of even a very complex mixture can be separated and analyzed individually.

By the use of paper chromatography, the amino acid analysis of a number of different proteins was carried

through. For instance, a molecule of the albumin of human blood plasma was found to be made up of the numbers of each of the nineteen amino acids listed in Table 6.

Table 6

THE AMINO ACIDS OF PLASMA ALBUMIN

Amino Acid	Number in Protein
Glutamic acid	80
Leucine	58
Lysine	58
Aspartic acid	46
Valine	45
Phenylalanine	33
Proline	31
Threonine	27
Arginine	25
Serine	22
Tyrosine	18
Cystine	16
Histidine	16
Glycine	15
Isoleucine	9
Methionine	6
Cysteine	4
Tryptophan	1
Alanine	0
<hr/>	
Total	510

Of course, knowing the mere number of each kind of amino acid in a particular protein molecule does not tell enough of the story. In what order do the amino acids appear?

This would seem to be a problem of impossible scope, one beyond the hope of solution. The number of possible arrangements even of small protein molecules is so unimaginably great that the selection of the one right arrangement out of the virtually infinite number of wrong ones would certainly seem the seeking after a needle in a universe of haystacks.

That the problem could nevertheless be solved is a tribute to human ingenuity as displayed in the person of the English biochemist Frederick Sanger and those who worked with

him. He dealt with insulin, a rather small molecule for a protein, being composed of only 50 amino acids and having a molecular weight of only 6000. (Even so, the number of possible arrangements of those amino acids was far greater than 10^{100} ; that is, a 1 followed by 100 zeros.)

Sanger's method, stripped to its essentials, was that of breaking down the molecule only part of the way, ending with peptide fragments made up of two, three, or four amino acids. The order of the amino acids in the fragments could be worked out and one could then reason backward to work out the order of larger fragments incorporating the smaller ones. Eventually it was possible to work backward to the one over-all order of amino acids that would yield all the fragments that had been detected, with amino acids in the proper order, and would serve as a source for no other fragments.

This is not as easy in practice as it might sound in words. The work involved was painstaking, time-consuming, and extremely delicate. Nevertheless, by 1953, after nearly a decade of research, the exact structure of the insulin molecule was worked out, amino acid by amino acid. It was then possible to show how insulin obtained from swine differed from insulin obtained from cattle, in the substitution of one amino acid for another in this position or that.

The same method was used to work out the structure of other and even more complicated protein molecules. By 1959 the structure of the enzyme "ribonuclease," with 121 amino acids in its molecule, was worked out in its entirety.

Once the order of the amino acids was known, it became possible to think seriously of tackling the vast entropy decrease involved in synthesizing a specific protein molecule by combining the amino acids in just the right way. This is no simple task at all. The protein molecule in question had to be built up by means of a number of steps, each purposefully directed toward the addition of one particular amino acid and no other as the next in line. Naturally, only the smallest protein molecules stood a hope of synthesis just at first. Thus the first molecule to be synthesized (by the American biochemist Vincent du Vigneaud and his coworkers, in 1953) was "oxytocin," a hormone produced by the pituitary gland. It possesses so small a molecule, made up of eight amino acids only, that it scarcely deserves the name of protein, yet the feat was greeted with great applause.

For one thing, the synthetic product was found to have all the properties of the natural one, so that synthesis proved the molecular structure of the hormone directly, where previously that structure had only been deduced, on the basis of more or less indirect evidence.

In 1960 another American biochemist, Klaus Hofmann, went a step further and succeeded in synthesizing a 23-amino acid portion of the molecule of "adrenocorticotrophic hormone," another product of the pituitary.

All in all, chemists are making remarkable headway in coming to understand the amino acid make-up of protein molecules and, considering that the real spurt began only fifteen years ago, hopes for the future are sanguine indeed.

And yet the reason for the flimsiness of the protein molecule has still not appeared in anything I have said so far.

Chapter 17

THE WEAK ATTRACTION

Can we be certain that the mere size of the peptide chain making up a protein molecule is not enough to insure the fragility of that molecule? In answer I can point out that mere size in itself is certainly not enough, since the cellulose molecule is as large and is not fragile. Nor can we attribute particular fragility to a polypeptide chain as opposed to other long chains, for there exist protein molecules, made up of chains of amino acids held together by peptide bonds, that are less fragile than cellulose molecules.

The best example of the latter is a protein called "fibroin," which has a rather simple make-up for a protein. Of the amino acids composing its molecule, five sixths are glycine, alanine, and serine, the simplest of all. Other amino acids are present in minor quantities and five are absent altogether. Yet fibroin is a protein, being made up of amino acids held together by peptide links. As the name implies, fibroin is most familiar to us in the form of a fiber. It is the chief

component of silk. And since a strand of silk is stronger than a strand of cotton (which is itself mostly cellulose), we must certainly assume there is nothing intrinsically unstable in the polypeptide chain itself.

And yet the fact remains that many protein molecules, including all enzymes, though made up of the same components and held together by the same types of bonds as is the fibroin molecule, are nevertheless changed irreversibly by gentle heating. Here is the germ of a paradox. The same components bound in the same manner can be as strong and tough as silk or as fragile as raw egg white.

Moreover, the change that converts raw, mucilaginous egg white into the solid white of the hard-boiled egg cannot involve the amino acids or the links between them; at least not the links I have been describing. The heat that is sufficient to solidify egg white or inactivate an enzyme is simply insufficient to break apart a peptide chain. A "denatured protein" (one that is no longer capable of carrying out its biologic function) must contain the same peptide chains made up of the same amino acids in the same order as does the original "native protein" from which it was formed by gentle heat.

And yet we cannot take the attitude that anything mystical, anything outside the rational laws of nature, can have taken place. If denatured egg white has properties that differ from native egg white, then some difference must exist in the structure of egg white. If the molecule of a protein is denatured and if none of the ordinary bonds between atoms are broken in the process, then there must be other weaker bonds that are essential to the specific structure of the molecule and that can be broken by the gentle environmental effects that can bring about denaturation.

That this is so is further indicated by the fact that it is sometimes possible to detect atom combinations in denatured proteins that are not detected in native proteins. The combination most often involved is the "thiol group," made up of a sulfur atom and a hydrogen atom (—S—H) (see Fig. 25). This combination is found in the amino acid cysteine, which is present in almost all proteins.

There are chemicals that will combine with the thiol group to give a precipitate or a soluble but characteristically-colored compound. (Chemicals that can be used to detect specific atoms or atom combinations by means of easily visi-

ble changes such as this are called "test reagents.") If cysteine is tied up as part of a peptide chain, its side-chain bearing the thiol group is still free. The same test reagents that react with pure cysteine will also react with a peptide chain containing cysteine.

Yet many protein molecules containing cysteine (as can be proved when cysteine shows up among the amino acids produced by the acid breakdown of the molecule) do not react with the test reagent. Once denatured, however, the protein *does* react with the test reagent.

The simplest explanation is that, in native protein, the peptide chain is so folded that the side-chains of some of the amino acids are directed into a hollow interior of the molecule. If the thiol groups of cysteine are so innerly-directed, the test reagent cannot make contact with it, and no reaction takes place. After denaturation, the peptide chain is opened up, the thiol group is exposed, and promptly reacts.

The process of denaturation breaks up the orderly pattern into which the peptide chains are folded. The chains unfold and take on a random arrangement. This increase in disorder involves a rather large increase in entropy. To reverse the process, forming an intricately folded native protein out of the randomly oriented denatured one, conversely involves a large decrease in entropy and, in general, such "renaturation" (except where the denaturation process has not gone very far) cannot be brought about. Roughly speaking, renaturation is the equivalent of un-hard-boiling an egg.

Now the folded peptide chain of the original native protein molecule must be held in place by some sort of chemical binding. This binding must be weak enough to be broken by the mild conditions that will denature proteins.

To find out the actual nature of these weak bonds, upon which we may have to rely for an explanation of protein fragility and of enzyme properties, I shall have to digress a bit and take up the matter of chemical bonds in general, repeating some of the material in Chapter 10—but going into greater detail.

From the very birth of the atomic theory, it was clear that the different atoms in a molecule had to be held together by some binding force. Moreover, the manner in which the force manifested itself seemed to be characteristic of the individual element. For instance, it was found that

a single hydrogen atom was never bound to more than one other atom; a single oxygen atom was never bound to more

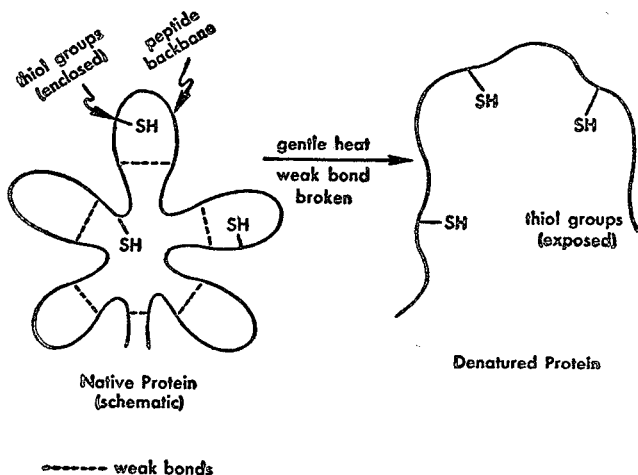


Figure 25. Denaturation of a Peptide Chain

than two other atoms; a single carbon atom was never bound to more than four other atoms, and so on.

The English chemist Edward Frankland, in 1852, suggested the name "valence" (from a Latin word meaning "power") for this combining power of the atoms of particular elements. Thus, hydrogen could be said to have a valence of 1, as have fluorine, chlorine, bromine, iodine, sodium, and potassium. Oxygen has a valence of 2 and so have sulfur, magnesium, and calcium. Nitrogen has a valence of 3 and so have phosphorus, arsenic, and aluminum. Carbon has a valence of 4 and so has silicon.

It is possible for an element to display more than one valence. For instance, although carbon almost invariably shows a valence of 4 it has a valence of 2 on rare occasions. Nitrogen usually has a valence of 3 but sometimes shows a valence of 5. Iron has valences of 2 and 3; gold, 1 and 3; copper and mercury, 1 and 2; and so on.

In connection with the relatively simple compounds of inorganic chemistry, the matter of valence did not affect the nature of the formulas already in use. Thus, iron combines with chlorine in two different ways. An atom of iron

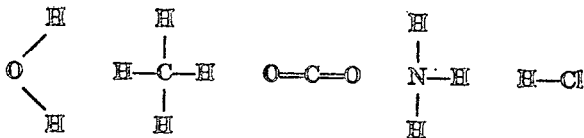
might combine with two atoms of chlorine to form "ferrous chloride" (FeCl_2), or with three to form "ferric chloride" (FeCl_3). The formulas used show this clearly and make it quite plain that iron has a valence of 2 in the first case and of 3 in the second.

Where the considerably more complicated compounds of organic chemistry are involved, however, the notion of valence proved a life-saver and revolutionized the system of writing formulas. The inadequacy of the ordinary formula in organic chemistry began to appear in the early nineteenth century when it was discovered that, among the organic chemicals, the molecules of different compounds did not necessarily have each its own characteristic atomic content.

The first to find this out, to his own astonishment, was a German chemist named Justus von Liebig, who in 1823 isolated a compound he called silver fulminate. He found that its molecule consisted of four atoms: one each of silver, carbon, nitrogen, and oxygen. The trouble with this was that a compound already known, silver isocyanate, had a molecule composed of those same four atoms and yet the properties of the two compounds were quite different.

Berzelius, when told of this, absolutely refused to believe it could be so until he himself located two other compounds of identical molecular composition yet different properties. Berzelius named such molecular twins "isomers" (Greek for "equal parts") and for a generation the problem of isomerism remained a mystery.

It was Frankland's systemization of valence that was the opening wedge to the solution, for in 1859 it occurred to the German chemist Friedrich August Kekule to represent each atom as possessing a number of hooks or "bonds" equal to its valence and to write those bonds into the formulas so that the atoms seemed held together Tinker-toy fashion. The formulas for water (H_2O), methane (CH_4), carbon dioxide (CO_2), ammonia (NH_3), and hydrogen chloride (HCl) would, according to this system, be written as follows:

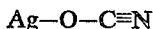


In these compounds, the hydrogen atom (H) and the chlorine atom (Cl) are always attached to one bond only, the oxygen atom (O) to two, the nitrogen atom (N) to three, and the carbon atom (C) to four. When the system is applied to more complicated organic compounds, as, for instance, sucrose or tryptophan, the system continues to work. The known atomic components of the molecule could always be made to fit together Tinker-toy fashion without violating the valence rules. Such Tinker-toy arrangements are called "structural formulas," whereas the mere listing of numbers of atoms, as in CH_4 , is an "empirical formula."

Almost at once it became clear that a fixed set of atoms might be put together in more than one fashion without violating the rules of valence. For instance, to take the first pair of isomers discovered, the formula for silver fulminate can be written as follows:



(Ag is the symbol for silver, which in Latin is "argentum.") The formula of silver isocyanate, on the other hand, is:



In either case, you see, the silver atom has the required valence of 1, oxygen 2, nitrogen 3, and carbon 4.

Where more complicated organic compounds are concerned, the number of possible arrangements that can be made without violating valence mounts rapidly and indeed becomes fantastically huge. It is calculated, for instance, that a molecule composed of forty carbon atoms and eighty-two hydrogen atoms can be arranged in more than sixty trillion ways.

For a century now, chemists have been able to use the Kekule system as a guide to the possible structure of new compounds and to the number of isomers possible in a given case. The system has been greatly refined and made at once more complicated and more flexible, but its main outline still stands and will stand, in all probability, for the foreseeable future.

To the chemists of the mid-nineteenth century, the chief problem concerning valence was that, however useful it was, it seemed arbitrary. Why should oxygen have a valence of 2, while carbon had a valence of 4, and hydrogen a valence of only 1? Why should hydrogen, sodium, and chlorine all be identical in valence, when they were so different in other ways?

The Russian chemist Dmitri Ivanovich Mendeleev weakened much of the force of this concern by showing valence to be less arbitrary than it seemed. In 1869 he listed the known elements in the order of their atomic weight and showed that their properties varied according to a fixed scheme. By placing the list of elements in rows and columns, he found he could arrange to have elements with very similar properties fall into vertical columns. (This is known as the "periodic table of the elements.") Furthermore, all the elements in a particular vertical column displayed the same valence (see Fig. 26).

As an example, fluorine, chlorine, bromine, and iodine all fell into the same column and all had a valence of 1; magnesium, calcium, strontium, and barium fell into the same column and all had a valence of 2; nitrogen, phosphorus, arsenic, and antimony all fell into the same column and all showed valences of 3 and 5; and so on.

Furthermore, as one went along the horizontal rows of the table, the valence changed in orderly fashion. For instance, at the beginning of the table the following elements are found, in order of increasing atomic weight, stretching across one row: lithium, beryllium, boron, carbon, nitrogen, oxygen, and fluorine. The valences are, respectively, 1, 2, 3, 4, 3 (or 5), 2, and 1.

The periodic table proved one of the great triumphs of chemical theory and valence was accepted as an orderly phenomenon even though the nature of the bond remained a mystery.

Then, about the turn of the century, the atom was found to have an internal structure itself. Its content of electrons was made plain. The atom, it turned out, consisted of a tiny, positively charged nucleus at the center, surrounded by a cloud of negatively charged electrons, filling the rest of the atomic volume. Each element is made up of atoms containing a nucleus with a characteristic charge. The hydrogen atom has a nucleus with the smallest possible charge,

Periods ↓	Groups →	I	II	III	IV	V	VI	VII	VIII	Period
1		1H							2He	1
2		3Li	4Be	5B	6C	7N	8O	9F	10Ne	2
3		11Na	12Mg	13Al	14Si	15P	16S	17Cl	18Ar	3
4		19K	20Ca	21Sc	22Ti	23V	24Cr	25Mn	26Fe	4
		29Cu	30Zn	31Ga	32Ge	33As	34Se	35Br	27Co	
									28Ni	
5		37Rb	38Sr	39Y	40Zr	41Nb	42Mo	43Tc	44Ru	5
		47Ag	48Cd	49In	50Sn	51Sb	52Te	53I	45Rh	
									46Pd	
6		55Cs	56Ba	57La	72Hf	73Ta	74W	75Re	76Os	6
		79Au	80Hg	81Tl	82Pb	83Bi	84Po	85At	77Ir	
									78Pt	
7		87Fr	88Ra	89Ac					86Rn	
6	$58-71\text{The lanthanides}$	58Ce	59Pr	60Nd	61Pm	62Sm	63Eu	64Gd	65Tb	6
7	$82-103\text{The actinides}$	90Th	91Pa	92U	93Np	94Pu	95Am	96Cm	97Bk	7
									98Cf	
									99Es	
									100Fm	
									101Mv	
									102No	
									103Lw	

Figure 26. Periodic Table of the Elements

which is arbitrarily considered to be 1. The carbon atom has a nucleus with a positive charge six times as great as that of the nucleus of a hydrogen atom, and its charge can therefore be written briefly and without ado as 6. By the same line of reasoning, the nitrogen nucleus has a charge of 7 and the oxygen nucleus a charge of 8. There is an element known at present for every charge from 1 to 103.

The free atom, ideally, contains a number of electrons (each with a negative charge of 1) that just balances the positive charge of the nucleus, so that the hydrogen atom has 1 electron, the carbon atom has 6, the nitrogen atom 7, the oxygen atom 8, and so on.

The electrons are arranged in "shells," each of which can contain some maximum number of electrons. The innermost shell can hold only two electrons, the shell outside that eight, the shell outside that eighteen, and so on. A carbon atom, with six electrons, must have them arranged in two shells, the innermost being filled with two so that the other four must be in the shell outside that. The chlorine atom with seventeen electrons must have them distributed among three shells; the innermost with two and the next with eight both being filled, the remaining seven must be in the third shell.

When two atoms collide, it is only the outermost shell of electrons of each atom that make contact, so to speak, and so the nature of the reactions undergone by these atoms depends on the number and arrangement of electrons in that outermost shell only.

This at once explains the similarities revealed by the periodic table. To give an example, the nine electrons of the fluorine atom are arranged in two shells, with 7 electrons in the second; the seventeen electrons of the chlorine atom are arranged in three shells, with 7 electrons in the third; the thirty-five electrons of bromine are arranged in four shells, with 7 electrons in the fourth; and the fifty-three electrons of iodine are arranged in five shells, with 7 electrons in the fifth. In each case the outermost shell contains 7 electrons, and all four elements are consequently similar in behavior.

As the twentieth century wore on and chemists grew more familiar with these facts, it became increasingly common to indicate the element by the ordinary symbol plus the num-

ber of electrons in the outermost shell only. For instance, hydrogen, which has only a single electron, would be $H \cdot$. Carbon, which has six electrons, but only four in the outer shell,

would be $\cdot \overset{\cdot}{\underset{\cdot}{C}} \cdot$. Table 7 lists some of the elements that are

important constituents of living tissue and their "electronic symbols."

Table 7
ELECTRONIC SYMBOLS

Element	Symbol
Hydrogen	$H \cdot$
Carbon	$\cdot \overset{\cdot}{\underset{\cdot}{C}} \cdot$
Nitrogen	$\cdot N \cdot$
Oxygen	$\cdot \overset{\cdot \cdot}{\underset{\cdot \cdot}{O}} \cdot$
Sodium	$Na \cdot$
Magnesium	$Mg \cdot$
Phosphorus	$\cdot P \cdot$
Sulfur	$\cdot \overset{\cdot \cdot}{\underset{\cdot \cdot}{S}} \cdot$
Chlorine	$\cdot Cl \cdot$
Potassium	$K \cdot$
Calcium	$Ca \cdot$

In the case of the lighter elements, such as those listed in Table 7, the most stable arrangement of electrons occurs when the outermost shell contains eight electrons. (An im-

portant exception is helium, which contains only one shell, that shell being able to hold only two electrons. A content of two electrons is the most stable situation in that case.)

Now consider the case of sodium. The eleven electrons of its atoms are arranged as follows: two in the innermost shell, eight in the next, and one in the third. If such an atom can manage to lose that one electron in the third shell, it will then have an outermost shell with a full complement of eight electrons, a stable situation. There is, for this reason, a strong tendency for a sodium atom to lose one electron.

As for the chlorine atom, its seventeen electrons are arranged two, eight, and seven. If such an atom can only find an additional electron somewhere, it could build up its outermost shell to the stable number of eight.

The needs of a sodium atom and a chlorine atom are thus complementary. If sodium and chlorine come into contact, an atom of the former will readily transfer an electron to an atom of the latter, with the result that both will be more stable than before. In losing an electron, the sodium atom is no longer able to balance the positive charge of the internal nucleus. The sodium nucleus has a positive charge of 11, but with one electron now missing, the ten electrons that remain can only counter that with a negative charge of 10. The net charge of the sodium atom, after loss of the electron, is thus $+1$. It has become the positively charged sodium ion, Na^+ .

On the other hand, the chlorine atom, with one additional electron, now has eighteen electrons with a total negative charge of 18, which overbalances the chlorine nucleus with its positive charge of 17. The net charge of the chlorine atom and its gained electron is now -1 . It has become the negatively charged chloride ion, Cl^- . The formation of ions, which had been worked out in the nineteenth century from a consideration of electric cells (see Chapter 10), was thus rationalized in the twentieth century out of a knowledge of atomic structure.

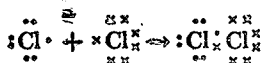
Once formed, the positively charged sodium ion and the negatively charged chloride ion are held together by the "electrostatic attraction" between the opposite charges, and the result is the compound sodium chloride, NaCl . A bond that results from electrostatic attraction, as between the so-

dium ions and chloride ions of sodium chloride, is called an "electrovalent bond."

It is thus clear why sodium has a valence of 1, since it has only one electron that it will easily donate before attaining a stable electronic arrangement. Again, chlorine naturally has a valence of 1, because it need accept only one electron to attain a stable electronic arrangement. The calcium and magnesium atoms have two electrons to donate out of their outermost shells and, naturally, have each a valence of 2. Oxygen, with six electrons in its outermost shell, has the capacity of accepting two electrons to attain stability and therefore has a valence of 2. If a calcium atom donates its two electrons to an oxygen atom, both atoms are stabilized to the calcium ion, Ca^{++} , and the oxide ion, $\text{O}^{=}$, and the strong electrovalent bond between them forms the stable compound, calcium oxide, CaO .

An interatomic bond need not always involve the transfer of one or more electrons from one atom to another. How, for instance, are the two chlorine atoms of a chlorine molecule held together? Each atom requires one additional electron and each has an equal tendency to gain one and no great tendency to lose one. In forming a two-atom molecule, neither atom can be expected to donate an electron to fill the other's need.

Instead, in true sisterly fashion, a pool of electrons is formed that may be shared equally by each. To show how this can take place, we can represent two atoms of chlorine with the electrons of one shown as dots and the electrons of the other as crosses. (This is purely a matter of didactic convenience and is not to be taken to indicate that the electrons of two different atoms are different in any way. As far as we know, all electrons of all atoms are identical.) The combination of two chlorine atoms to form a chlorine molecule can then be written as follows:



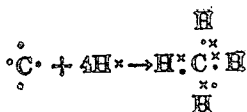
The two chlorine atoms of the chlorine molecule possess a total of fourteen outer-shell electrons all together; no new electrons have been created. However, the two electrons rep-

resented as lying between the atoms (one originating from each atom, as the cross and dot show) are within the outer shell of each atom. Both atoms, therefore, have an outer shell containing eight electrons and the arrangement as a two-atom molecule is far more stable than the arrangement as two separate atoms. Since the two shared electrons can form part of the outer shell of both atoms only if the atoms remain close together, it takes energy to pull the atoms apart and pass from a more stable to a less stable arrangement.

In other words, the formation of molecular chlorine from atomic chlorine is accompanied by a decrease in free energy and is a spontaneous reaction. The dissociation of the chlorine molecule involves an increase in free energy and will not take place spontaneously.

The presence of a shared pool of electrons thus acts as a bond holding atoms together. Since the two atoms are co-owners, so to speak, of the shared electrons, this type of bond is a "covalent bond." Most of the bonds in organic compounds are of the covalent variety.

Methane (CH_4) may be looked upon, electronically, as a combination of a carbon atom, with its four outer electrons, with four hydrogen atoms, each with its lone electron.



The carbon atom shares each of its electrons with a different hydrogen atom and each hydrogen atom contributes its electron to the shared pool. The result is that each hydrogen atom possesses two electrons in its only shell (all that shell will hold and therefore a stable arrangement) while the carbon atom possesses a stable total of eight.

By properly sharing the outer electrons (also called "valence electrons") all the facts previously interpreted by the original Kekule straight-line bonds could be reinterpreted adequately in electronic terms. In addition, some facts could be explained by electrons which the original Kekule system of writing formulas left unexplained.

Thus, consider that the two chlorine atoms of the chlorine molecule, being identical, must have an equal hold on elec-

trons. The two electrons of the shared pool between them are equally shared. Carbon and hydrogen atoms have an approximately equal hold on valence electrons, so that in methane the shared pairs of electrons are shared nearly equally.

But what about the covalent bond between oxygen and hydrogen? An oxygen atom will form a two-electron shared pool with a hydrogen atom, but the oxygen atom, having the stronger hold on valence electrons, will hold them more firmly, so to speak, in its own outermost shell. The electrons will occupy the electron shell of the hydrogen atom to a correspondingly lesser degree.

Since the oxygen atom has more than an equal share of the pooled electrons, it also has more than an equal share of the negative charge of those electrons. There will therefore be a fractional negative charge on the oxygen atom. The hydrogen atom, deprived of its fair share, will have a small (and just balancing) positive charge.

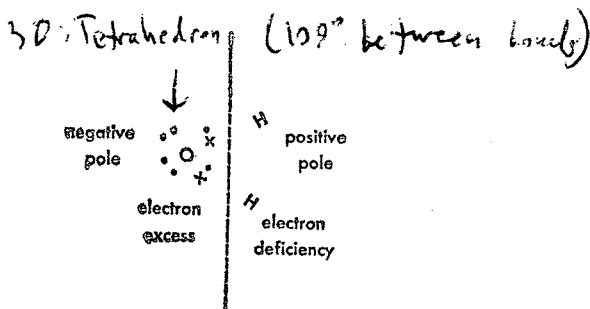


Figure 27. *A Polar Molecule of Water*

A molecule such as that of water will therefore have a small concentration of positive charge at the hydrogen end and a small concentration of negative charge at the oxygen end, as is shown in Fig. 27. These are like the two opposing poles of a magnet and can be referred to as poles of electric charge. For this reason, water is a "polar molecule." Molecules such as those of chlorine and methane, in which there are no permanent concentrations of charge at specific places, are "non-polar molecules."

Opposing electric charges on two polar molecules will attract each other, just as oppositely charged ions will. In

other words, the water molecules in liquid water (and much more so in ice) will have a tendency to arrange themselves so that the negatively charged oxygen end of one molecule is adjacent to the positively charged hydrogen end of its neighbor.

The charges involved in polar molecules are much smaller than the charges involved where electrons are completely transferred from one atom to another instead of being merely unequally shared. The attractions between polar molecules are not nearly as strong, therefore, as that of the electrovalent bond. These intermolecular attractions are, in fact, only about a twentieth as strong as the covalent bond.

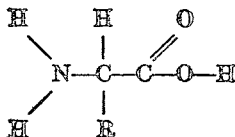
Nevertheless, these attractions between polar molecules are sufficiently strong so that water molecules, for instance, are noticeably "sticky." Far more energy is required to pull the sticky molecules of water apart (melting the solid or boiling the liquid) than is the case for non-sticky molecules such as those of methane.

Thus, although methane and water have molecules of roughly the same size and complexity, the melting points of methane and water are -184° C. and 0° C. respectively, while the boiling points are -161° C. and 100° C. respectively.

The intermolecular forces are strongest when a positively charged hydrogen atom makes up part of the molecule, because the hydrogen atom is so small (the smallest of all atoms) that strongly negatively charged atoms (such as those of oxygen, nitrogen, and fluorine) can make an unusually close approach. Since electrostatic attraction varies inversely as the square of the distance from atomic center to atomic center, an unusually close approach means an unusually strong attraction.

The intermolecular force is strong enough to be noticed, therefore, only in those cases where hydrogen atoms lie between two negatively charged atoms. In water, for example, the hydrogen lies between two oxygen atoms. Because of the central role of the hydrogen atom, this attractive force is called a "hydrogen bond" (see Fig. 28).

There is a clear application here to the problem of protein structure. To show this, let's begin by taking a look at the details of the amino acid molecule according to the Kekule system:



The NH_2 combination on the left is the amine group; the COOH combination on the right is the carboxylic acid group. The "R" attached to the central carbon atom represents the side-chain, this being different for each amino acid.

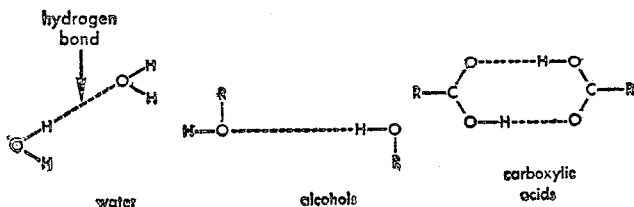


Figure 28. *The Hydrogen Bond*

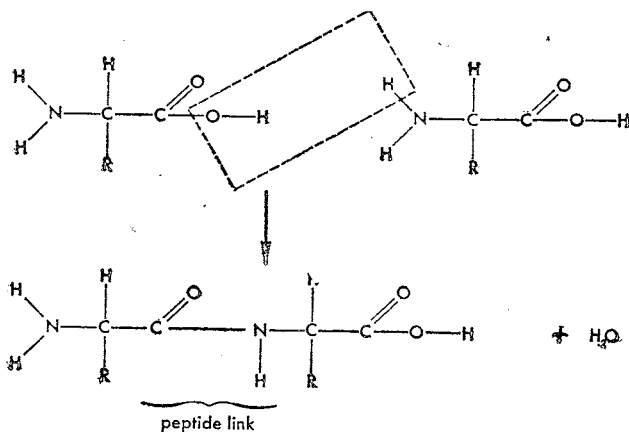
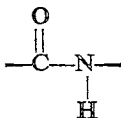


Figure 29. *Amino Acid Combination*

When two amino acids combine, an OH is removed from the carboxyl group of one and an H from the amine group of the other. These unite to form water. What is left of the amino acids combines at the bond sites left free by the departing OH and H. This is best shown by the equation in Fig. 29.

The peptide bond that links the amino acid components has, as you see, the structure:



If the amino acid chain is imagined to be doubling back upon itself, there can easily arise a situation where the hydrogen attached to the nitrogen of one peptide link would find itself in the neighborhood of the oxygen attached to the carbon of another peptide link. The hydrogen would

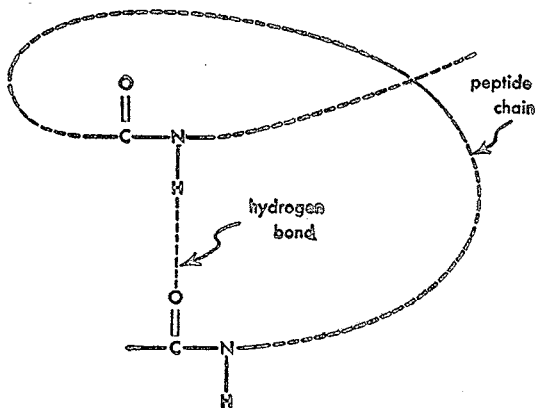


Figure 30. *Hydrogen Bonds and Peptide Chains*

then lie between the oxygen and the nitrogen and the requirements for a hydrogen bond would be fulfilled. The situation would appear something as in Fig. 30 and, as you see, this would just represent the situation shown earlier on

page 221. The hydrogen bond is the "weak bond" to which I there referred.

The American biochemists Linus Pauling and A. E. Mirsky suggested in 1936 that the presence of hydrogen bonds in the protein molecule would account for its fragility. The specific activity of such a molecule would depend, in this view, not merely on the existence of an amino acid chain in which the amino acids were present in one particular order. The activity would depend also on the chain being folded in a particular fashion, those folds being held in place by hydrogen bonds.

Mild heating would have no effect on the strong covalent bonds making up the peptide links but it would disrupt the hydrogen bonds, which are only a twentieth as strong. The folds of the protein molecule would fall apart into a randomly looped and coiled amino acid chain and the result would be a denatured protein, one in which the specific activity had been lost.

But in order to determine exactly how an amino acid chain might fold, it was necessary first to know the exact geometry of a molecule. How far apart are the neighboring atoms of a molecule and what relative positions do they occupy? In other words, what are the lengths of the bonds between atoms and what angles does one bond make with another? (Although a structural formula is written two-dimensionally on paper, please remember that in actual fact the atoms of a molecule are distributed through three dimensions.)

The answer to these questions came through the use of X rays. A beam of X rays impinging on a crystal of some substance is partially reflected by the ordered layers of atoms that characteristically make up a crystal. From an analysis of the manner in which the X rays are reflected, it is possible to deduce the distances between layers of similar atoms within the crystals. From this one can in turn deduce the pattern of the crystal structure.

This method depends on the existence of an ordered atomic arrangement within the object being subjected to the X ray beam. In the early 1920s, the X ray studies were made of silk fibers. These are not crystals, of course, but there is enough regularity in the amino acid chains within silk to produce significant X ray reflection. From the patterns produced, it was possible to show, for instance, that a single

amino acid component of the peptide chain is 0.35 millimicrons (0.000000014 inches) long. Individual bond lengths and bond angles could then be calculated. The same technique was applied to other proteins, such as the keratin occurring in hair, in the 1930s.

In the early 1950s, Pauling and his coworkers, using such data, showed that it was possible for peptide chains to coil into helices (shapes similar to that of a spiral staircase) without putting any strain on the natural orientation of the various bonds. Helices could be designed in such a fashion that hydrogen bonds just naturally fell into place here and

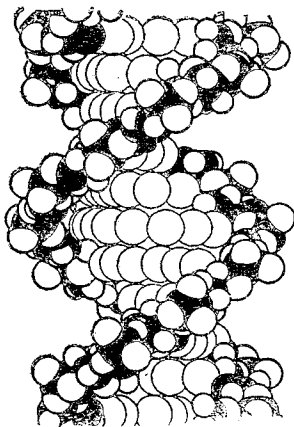


Figure 31. *A Helical Molecule*

there on the coil. Furthermore, several such coils could combine into a more elaborate coil (precisely as a rope is made up of twisting fibers that are in turn made up of twisting fibers), the neighboring coils again held together by hydrogen bonds (see Fig. 31).

Experimental evidence has backed this view, and currently the notion of a helical protein structure is accepted by most biochemists.

Chapter 18

THE WORKINGS OF ENZYMES

The helical structure of the protein molecule, held together, as it is, by hydrogen bonds, explains not only the fragility of an enzyme and the ease with which it is inactivated, but also its great specificity.

Let's compare powdered platinum and an enzyme molecule. As I explained in Chapter 12, platinum probably owes its catalytic activity to its ability to bind to its surface by spare valence bonds those molecules whose reaction it catalyzes. Such binding is most likely to take place at those areas on the surface of the platinum particle where there are irregularities. A molecule with a particular configuration of atoms might find itself just fitting snugly over the hump of a particular clump of platinum atoms, or in the valley between two clumps.

Since the platinum surface undoubtedly contains all sorts and shapes of irregularities there would be something for almost any molecule. The result is that a wide variety of molecules would find themselves capable of being temporarily locked to the platinum surface and able to react more readily. For that reason platinum and other catalysts of its type are comparatively unspecific in the reactions they will catalyze. Such catalysts are also relatively inefficient, for only a small percentage of the irregularities might suit the molecules involved in a particular reaction; and for any particular reaction, most of the surface would remain non-catalytic in nature.

The enzyme molecule is quite another thing. It offers no random surface but, instead, a specific surface constructed by the chemical processes of the organism (see Fig. 32). The varying side-chains of the individual amino acids composing the peptide chain represent irregularities of different sizes and types. Some side-chains, like those of aspartic acid and

specific

glutamic acid, possess negative charges that would attract any positive charge that might exist on molecules in the neighborhood. Others, like those of lysine, arginine, and histidine, possess positive charges that would attract a negative

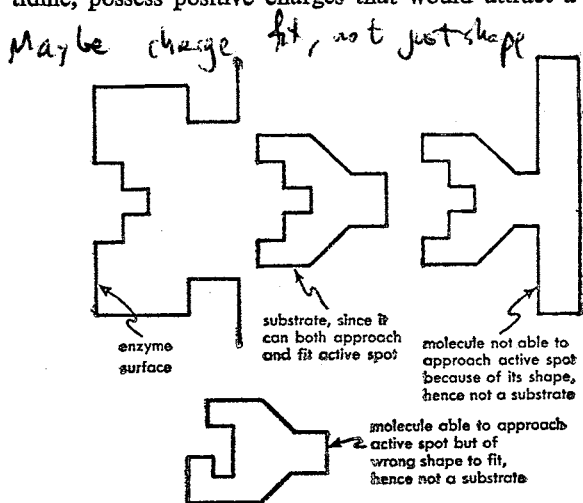


Figure 32. Enzyme and Substrate (Schematic)

charge. Still others, like serine, threonine, and tyrosine, lack charges but possess groups capable of forming a hydrogen bond.

As for side-chains that do not form bonds, such as those of valine, alanine, leucine, isoleucine, phenylalanine, and a few others, those act as barriers of various forms so that only molecules of a given shape can fit among the barriers and approach those areas of the enzyme surface that can properly grip it.

All these side-chains are held in fixed relationship to each other by the hydrogen bonds that keep the protein coil in shape. The result is that only a molecule of a particular type can actually bind to a given enzyme, and it is only that molecule or type of molecule that will undergo a reaction catalyzed by that particular enzyme. (The molecule that undergoes an enzyme-catalyzed reaction is termed the "substrate" of that enzyme.) And, of course, anything that disrupts the hydrogen bonds and upsets the fixed relationship

of the side-chains destroys the special affinity of enzyme for substrate and ends the enzyme's catalytic activity.

This tight specificity of an enzyme molecule has an interesting consequence. Any consideration of the chemical machinery of life shows that the functioning of an organism is dependent on the smooth workings of many thousands of different reactions, involving many thousands of different substances. With enzymes as specific as they are, would that not mean that each organism, and even each cell of each organism, must be supplied with thousands of different enzymes, each intended for one particular reaction?

This is, indeed, exactly the state of affairs. Although the cell is a small object, perhaps 0.03 millimeters in diameter, it is tremendously larger than a protein molecule. There is room in the cell for thousands of different enzymes. What's more, there is room for thousands of molecules of each of these different enzymes.

This may seem, at first thought, a wasteful and inefficient situation. Actually it is quite the reverse; it is a system that, for economy and efficiency, is beyond praise.

Each enzyme, specifically designed for its purpose, has a far larger proportion of "active spot" on its surface than is to be found on the randomly shaped surface of a platinum particle. It is therefore far more efficient.

Thus the rate at which a reaction is brought about by a catalyst can be measured as the "turnover number." This is defined as the number of molecules of substrate undergoing a reaction each minute, at a particular temperature, under the influence of a single molecule of enzyme. The enzyme "catalase," which brings about the breakdown of hydrogen peroxide to water and oxygen, has a turnover number of 2,500,000 at 0° C. Most enzymes have turnover numbers that are considerably lower, to be sure; nevertheless these values are much higher than the corresponding values for ordinary non-enzymatic catalysts.

Supposing the various enzymes to be arranged systematically within the cell (and they are, as we shall eventually see), it is possible to envisage substrates moving along from place to place being systematically changed in microseconds.

In addition, by having a vast number of specialized catalysts, rather than a few general ones, the organism is supplied with a method of controlling cell chemistry with the utmost delicacy. By increasing or decreasing the number of

molecules of this enzyme or that within the cell, a particular reaction out of all the thousands simultaneously proceeding may be accelerated or decelerated, with, in each case, a subtly different over-all result.

This picture of an organism being composed of many thousands of intermeshing chemical reactions under the separate control of many thousands of enzymes makes it inevitable that even a primitive life form should be more complex than man's most intricate devices. Enlarge the machinery of life from the molecular level at which it exists to the size of the mechanical components of man-made instruments, or even only to the size of the miniaturized electronic components of modern computers, and a single cell would grow to tremendous size.

The fact of the matter is that the living cell is made up of a far greater number of moving parts (so to speak) than anything man-created, and it is only the fantastically small size of those parts that gives us the illusion that the cell is "nothing more" than a microscopic blob of jelly. As for organisms composed of trillions of cells, the increase in complexity outstrips the powers of description. To compare the lizard with the rock, except in terms of gross size, is laughable. For that matter, I would not scruple to compare the complexity of a man with that of a star—in favor of man.

A star isn't very complex, is it?

Of course, this analysis of enzyme specificity is based on the assumption that the substrate makes physical contact with the enzyme and forms an "enzyme-substrate complex" as an intermediate. It is difficult not to make this assumption because for an enzyme to have an effect on a molecule, without touching it somehow, is a mystical notion of a type scientists avoid.

Furthermore, early evidence in favor of such a complex, although indirect, was quite convincing. For instance, suppose a little substrate is added to an enzyme solution. A reaction begins at once and chemists can usually design methods for determining the rate at which substrate disappears or at which a product of the reaction appears. This gives them the "reaction rate." If substrate had been added in greater quantity to that same enzyme solution, the reaction rate would have been higher; just as, to use an analogy, the rate of outflow of goods from a supermarket

is increased when a larger number of customers enter and shop in the store. In fact, if the number of customers in the supermarket is suddenly doubled, you would expect the quantity of goods leaving the store, per unit time, to double. And, analogously, if the substrate concentration is doubled, the rate of reaction would be doubled.

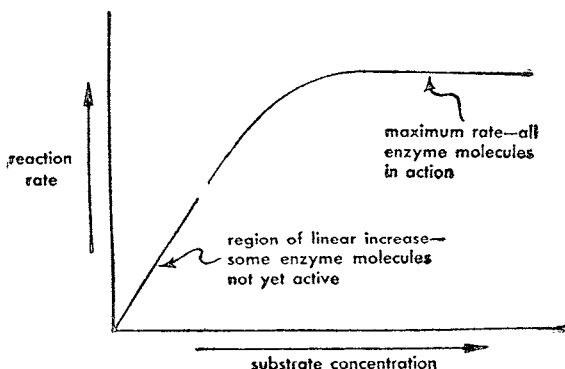


Figure 33. *Reaction Rate and Substrate Concentration*

This, however, continues only up to a point. Once the substrate concentration reaches a certain level, the reaction rate hits a maximum and increases no more, no matter how much substrate is added (see Fig. 33). The supermarket analogy holds here, too, for the indefinite piling up of customer concentration does not increase the goods-outflow rate indefinitely. Once all the checkout counters are in full use, a further increase in customers only lengthens the queue at each counter.

In the same way, once the substrate concentration is such that each enzyme molecule is utilizing its active center at maximum speed, a further increase of substrate only lengthens the molecular queue, waiting, so to speak, for one of the centers to be available. The assumption that enzyme and substrate form a temporary complex is the simplest way of explaining this observed phenomenon, a point first clearly made in 1902 by a French biochemist, V. Henri.

Two German biochemists, Leonor Michaelis and M. L. Menten, in 1913 took this assumption and supplied it with a mathematical treatment similar to that used by chemists

for ordinary reaction rates. They ended with an equation that represented the change of reaction rate with increasing substrate concentration to be just as actual observation indicated it to be. By observing the actual change in particular cases, they could make use of their equations to calculate the force holding a particular substrate to a particular enzyme. The "Michaelis-Menten equation" has been used for this purpose ever since.

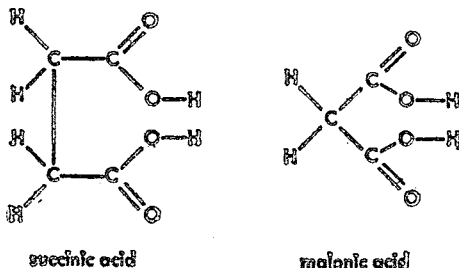


Figure 34. *Molecular Competitors*

Now if a particular assumption gives rise to a mathematical relationship that seems to express, accurately, an actual physical phenomenon, this is not necessarily proof of the truth of the assumption, but it is certainly evidence in favor of the truth.

Additional evidence came from another direction. Suppose a substrate is quite similar in structure to another compound. What happens if the other compound is added to the enzyme solution in place of the substrate? This situation arose in 1930, when the biochemist J. H. Quastel was working with an enzyme for which "succinic acid" is the substrate (see Fig. 34).

The structural formula of succinic acid is very similar to that of another substance, "malonic acid" (see Fig. 34). If malonic acid is added to the enzyme in place of succinic acid, no reaction takes place. Malonic acid lacks only a CH_2 group, and is otherwise similar to succinic acid, yet the enzyme can clearly distinguish between them.

Yet if malonic acid is added, the enzyme solution is not completely oblivious of the fact. If, with the malonic acid still there, succinic acid, the normal substrate, is *then* added, there is no reaction. The malonic acid has "poisoned" the

enzyme, or, to use a less dramatic term, it has "inhibited" the enzyme.

The easy deduction to be made is that malonic acid has a molecule sufficiently similar to that of succinic acid to jam itself into position on the enzyme's active spot, but is sufficiently different so that nothing further happens. Furthermore, its binding is so strong that it is not released. It

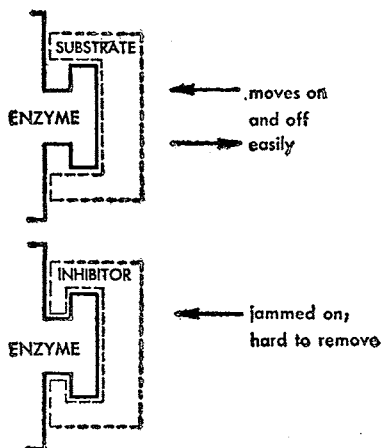


Figure 35. *Competitive Inhibition (Schematic)*

is like the slightly wrong key jammed into the lock and broken off there in a futile wrench. It does not open the door, and now the right key will not open it either. (In fact, the theory of enzyme action that has a substrate just fitting a spot on the enzyme surface is called the "lock-and-key mechanism." See Fig. 35.)

Because malonic acid (the "inhibitor") seems to compete with the rightful succinic acid for a position on the enzyme's surface, this process is called "competitive inhibition." Since 1930 a large number of examples of competitive inhibition have been studied. The Michaelis-Menten equation has even been modified to allow for the formation of enzyme-inhibitor complexes as well as enzyme-substrate complexes, and again, the mathematical deductions have accurately represented the observed phenomena.

Competitive inhibition offers a method for controlling the

rate of enzyme-catalyzed reactions by some means short of actually removing some of the enzyme or of making more of it. There are a number of groups of compounds of importance to the body that are quite similar in structure among themselves. The amino acids valine, leucine, and isoleucine are similar. The sugars glucose and galactose are similar, and so on. Competitive inhibition among them is almost inevitable. It is likely that the presence of such compounds in varying proportions within cells competitively inhibits this enzyme or that to a degree depending on the exact relative concentrations and that this, in effect, delicately swings the chemistry of the cell in a desired direction, or, for that matter, sensitively holds it in line. (This is a kind of automation on the molecular level.)

Even more dramatic is the competitive inhibition brought about by the deliberate addition of compounds to an organism; compounds, in some cases, that are of laboratory origin and that are complete strangers to living tissue. Such deliberately induced competitive inhibition will allow a distinction to be made between one organism and another even when the two are in intimate contact, as is the case with a bacterial parasite and the host it infests.

Many poisons achieve their results by drastically interfering with the activity of many enzymes. Such a poison as, for instance, mercury bichloride ("corrosive sublimate") will kill all forms of life, and to kill germ and patient together is little satisfaction.

By using a competitive inhibitor, however, one and only one enzyme can be put out of action. By a proper choice of dosage, the bacterial enzyme may be interfered with while the human enzyme may remain largely unaffected, either because the bacterial enzyme is more sensitive than the human enzyme, or because the membrane of the bacterial cell will allow the inhibitor to penetrate more quickly than will the membrane of the human cell. It may even be that the particular enzyme is more essential to the bacterial chemical machinery than to that of the human.

The first important case in point was that of "sulfanilamide" (see Fig. 36), a compound first synthesized in 1908. In 1932 a German biochemist was testing various dyes to see which, if any, could kill bacteria without serious damage to higher organisms. One, called "Prontosil," was found very

effective against certain streptococci and the news was released to the world in 1934.

However, Prontosil, while it worked if injected into an infected mouse, did not work against the bacteria in a test tube. Presumably, then, it wasn't Prontosil that did the work, but some substance formed from Prontosil by the living organism. French biochemists broke down the Prontosil molecule and obtained a component portion of it that turned out to be sulfanilamide. This proved active against certain

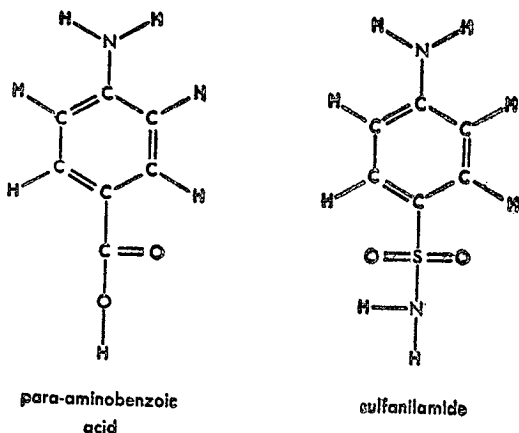


Figure 36. *More Molecular Competitors*

bacteria both in the organism and in the test tube.

This was the first of a whole series of "miracle drugs" which, in a generation, virtually completed the victory of man over many an infectious disease that had plagued him through all his history.

As it turns out, sulfanilamide is similar in structure to a compound called "para-aminobenzoic acid" (see Fig. 36), which forms part of an essential component of all cells. Clearly, sulfanilamide is competitively inhibiting a bacterial enzyme dealing with para-aminobenzoic acid and, in so doing, kills the bacteria.

Since then, the various antibiotics (penicillin, streptomycin, and so on) have been introduced and it seems almost certain that each one works through one form or another

of competitive inhibition, though biochemists are not certain as yet exactly where such blocking actions may be exerted.

The lock-and-key mechanism whereby an enzyme can distinguish between two compounds of different structure seems most remarkable in connection with "optically active compounds." The use of this term, however, requires a certain amount of explanation and for this we must begin with light.

From ancient times it was known that a ray of light, passing from air into some transparent medium, such as water, was bent or "refracted." In 1670, however, a Danish physicist, Erasmus Bartholin, discovered that a crystal of Iceland spar (a variety of calcium carbonate, CaCO_3) refracted a ray of light in two different ways, forming two rays of light where one existed to begin with. This was called "double refraction."

Double refraction was a stumbling block to all theories of the nature of light for more than a century. However, at the beginning of the nineteenth century it was established that light behaved as though it consisted of tiny waves and, with that, the matter of double refraction was soon cleared up.

In an ordinary ray of light, such as one from the sun or from any heated object, the light waves undulate in all possible planes. Some undulate up and down, some right and left, and some at all intermediate positions. In Iceland spar, however, the regular arrangement of particles in the crystal (and the atomic theory was also being established at the beginning of the nineteenth century) is such that it forces the waves of light to undulate in one of only two directions, one at right angles to the other.

Two types of light rays are therefore formed of the light that enters the crystal: one in which the waves all vibrate up and down (let us say) and one in which they all vibrate left and right. The two light rays have different properties and are refracted by different amounts, so that where one ray enters the crystal two rays leave.

Any ray of light with waves vibrating in a single plane is called "plane-polarized light," or simply "polarized light." This term (a poor one) was introduced by a French engineer, Etienne Louis Malus, who discovered in 1810 that light re-

flected from glass at a particular angle was also set to undulating in a single plane.

Techniques were worked out for detecting the exact plane in which a particular ray of polarized light undulated, and the best of these was developed by the British physicist William Nicol in 1828. He put together two prisms of Iceland spar and cemented them with Canada balsam in such a way that one of the polarized rays passed through, while the other was reflected away altogether at the line of junc-

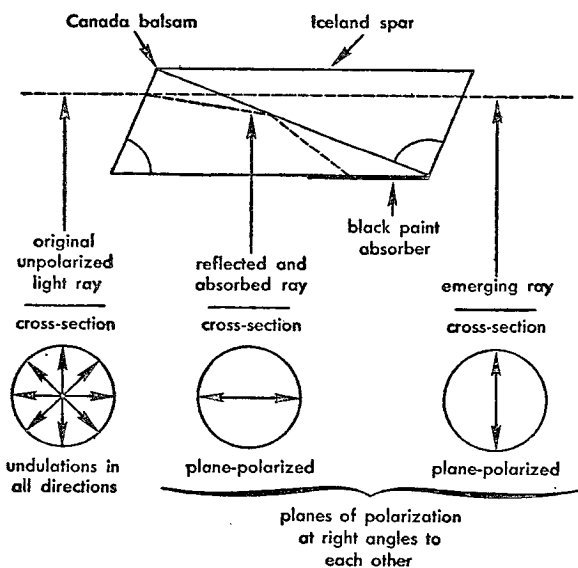


Figure 37. Nicol Prism

tion. Only one ray, a polarized one, emerged (see Fig. 37).

The emerging ray can pass through another such "Nicol prism" if the second is oriented just parallel to the first. If the second Nicol prism is slowly rotated, less and less of the polarized ray can get through until, when the second Nicol prism is set at right angles to the first, none of the light can get through. A device incorporating two such prisms is a "polarimeter." (Nowadays, in place of the bulky and expensive Nicol prisms, sheets of plastic in which cer-

tain organic crystals have been oriented are used for the job. A familiar trade name for such a substance is Polaroid.)

If the ray of polarized light travels through air or water in going from one Nicol prism to the second, nothing happens to the plane in which it undulates. Light passes through the second prism, held in parallel orientation, without significant diminution.

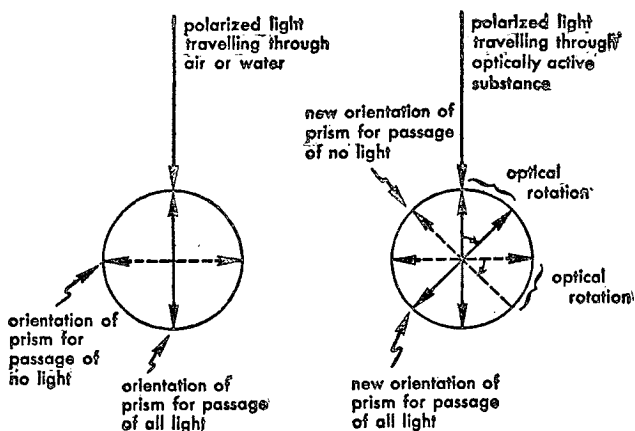


Figure 38. *Optical Activity*

A French physicist, Jean Baptiste Biot, in 1815 had discovered, however, that if polarized light passes through quartz crystals the plane of vibration is changed. With the development of the polarimeter, it could be easily shown that the light passing through the quartz crystal would then grow appreciably dimmer when it passed through the Nicol prism. The second Nicol prism had to be rotated until its orientation was parallel to the new plane into which the quartz had twisted the undulations of the polarized light passing through it. The degree to which the plane of polarization had been twisted could be measured with considerable precision by noting the change in orientation of the second Nicol prism that became necessary in order to restore the transmitted light to its full brilliance (see Fig. 38).

Biot also showed that certain organic compounds in

solution—turpentine and camphor, for example—could twist the plane of polarized light. This behavior is termed “optical activity.” Biot maintained that there must be some asymmetry in such optically active substances, asymmetry that produced a twist upon the ray of light, always in the same direction. Symmetrical substances, he felt, would also have the tendency to twist but in both directions equally, leaving no over-all twist at all.

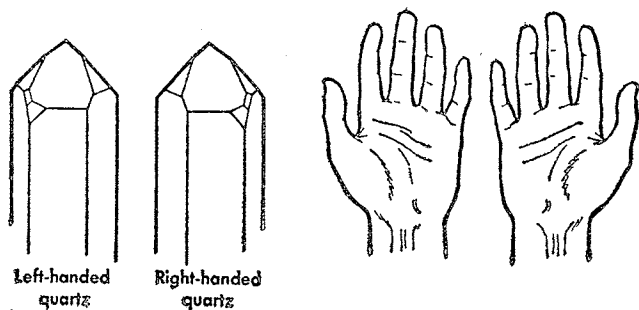


Figure 39. *Bilateral Asymmetry*

In 1848 a young French chemist, Louis Pasteur (then only twenty-six and at the beginning of one of the most brilliant scientific careers in history), electrified the world of chemistry by showing this was literally so, at least in the case of crystals (see Fig. 39). Working with crystals of an organic substance called sodium ammonium tartrate, he found that he could isolate two forms of the crystal, both forms asymmetric and one form being the mirror image of the other. (Your hands and your feet are asymmetric in this fashion.)

Carefully, Pasteur used tweezers and a magnifying glass to separate the two types of crystals into separate heaps and found that in solution both were optically active. What's more, one twisted the plane of polarized light clockwise and the other twisted the plane counterclockwise. If the two solutions were equally mixed, there was no twisting at all.

The separation of the two forms of crystals, changing a random mixture of the two into separate heaps containing one each, represents an increase of order and, therefore, a decrease of entropy. It is analogous to separate carefully a large deck of cards, randomly shuffled, into two piles, one

of red suits and one of black. Here is an excellent example of how such a decrease is at the expense of a larger increase of entropy. In order for the crystals to be separated, Pasteur had to pore over his magnifying glass and poke with careful persistence among the crystals with his tweezers. One can well imagine that the increase in *his* entropy in the process was far greater than the decrease he brought about among the crystals.

In any case, Pasteur succeeded in this fashion in demonstrating asymmetry only for crystals. In solution, however, optically active substances existed as single molecules. Where was the asymmetry, then? In the molecules themselves?

The answer was, Yes! It came in 1874, when two young chemists, the Dutchman Jacobus Hendricus van't Hoff (age twenty-two) and the Frenchman Joseph Achille Le Bel (age twenty-seven), independently presented a new theory of the carbon atom that could refine and expand the Kekule system of formulas to account for some facts for which the Kekule system could not account.

When structural formulas are drawn according to the Kekule method, the carbon bonds are drawn out to the four points of the compass, as, indeed, is the natural way to do it on a two-dimensional surface. However, this doesn't actually represent the true facts of the case, for if such a picture were accurate, there should be two forms of the compound dichloromethane (CH_2Cl_2), corresponding to the formulas shown here:



where, in one case, the chlorine atoms are adjacent and in the other are not. Yet in actual fact only one form of the compound has ever been detected.

The Van't Hoff-Le Bel theory was that the four bonds of the carbon atom existed in two planes at right angles to each other (see Fig. 40). This is hard to show in a two-dimensional drawing but the simplest way is to imagine three of the bonds forming a squat tripod while the fourth sticks straight up in the air. The four bonds point to the corners

of an imaginary regular tetrahedron and the angle between any two bonds is equal, and about 109° . The carbon atom could be set down in any fashion and the three bonds below would form the same sort of tripod, while the fourth would stick straight up.

Since each bond is equidistant from the three others, then if dichloromethane is formed according to this system, it doesn't matter which bonds are used for the two chlorine atoms. All arrangements are identical. You can see this if

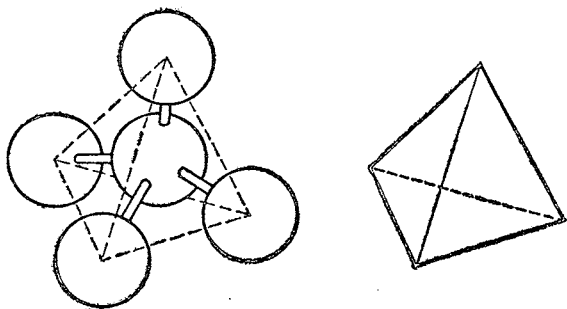
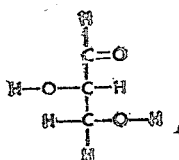


Figure 40. *The Tetrahedral Carbon Atom*

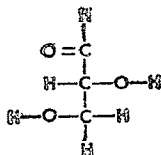
you set up two models made of a central plasticine or wax ball into which sticks are pushed, in the tripod arrangement I have already described. If you use two white sticks (hydrogen) and two green sticks (chlorine), then if you place them in various arrangements, you will find that one can always be turned into another if the ball is turned or twisted properly, without pulling out and shifting the position of the sticks themselves. Therefore, only one form of dichloromethane is possible by the Van't Hoff-Le Bel system, which is in accord with observation.

However, when a different group is attached to each of the four bonds of the carbon atom, the result is different. You can try the experiment yourself, using the wax ball and four sticks of different colors. You will find that all possible arrangements can be twisted into one of two arrangements, of which one is the mirror image of the other. You will find that no amount of twisting and turning will make one

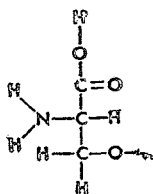
of these mirror images take up the exact pattern of the other. These two mirror images can be represented in the Kekule system as shown in Fig. 41 and one formula cannot be changed into the other by motions which (it is agreed) are confined to the plane of the paper.



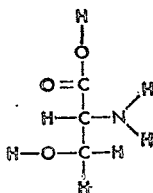
L-Glyceraldehyde



D-Glyceraldehyde



L-Serine



D-Serine

Figure 41. *D and L Series*

It was quickly found that in every case where an organic compound showed optical activity the molecule displayed asymmetry, due almost always to the presence of at least one "asymmetric carbon atom" (one to which four different groups were attached). One of the possible forms that could be written by the new system twisted light clockwise, and the other twisted light to exactly the same degree but counterclockwise. But which form twisted light in which way?

There seemed no possible way of deciding that, so the German chemist Emil Fischer decided in 1891 to make an arbitrary choice. He wrote the two forms of a compound called "glyceraldehyde" (which could be viewed as a very simple sugar molecule) and labeled one "D-glyceraldehyde"

and the other "L-glyceraldehyde." All other optically active compounds could be related to one or the other and together they belonged to the D-series or the L-series (see Fig. 41). (Recently it has been shown that Fischer guessed right. The compounds he placed in the two series actually do have the structure he picked for them.)

In the laboratory, when a compound with an asymmetric carbon atom is formed, both mirror images are formed in equal quantities, since that is the situation of maximum disorder and, therefore, of maximum entropy. To introduce more of one than another requires special devices, such as the utilization of a compound, in the process of manufacture, that is already optically active.

Such compounds can be obtained from plant and animal sources. In fact, all compounds in living tissue that have asymmetric carbon atoms exist there in one optically active form or the other. Tissues can decrease entropy rather admirably in this respect as they can in forming specific proteins—at the expense, of course, of an over-all increase in entropy.

Thus, it turned out that all the natural sugars belong to the D-series, and all the natural amino acids (such as serine, which is used as an illustration in Fig. 41) belong to the L-series. The body could not, in general, use L-series sugars or D-series amino acids and these are sometimes called "unnatural sugars" or "unnatural amino acids." Actually they do occur in nature to a small extent. There is an L-sugar in streptomycin and D-amino acids in certain antibiotics. A compound, called "ascorbic acid," that is vital to life is related to the "unnatural" L-sugars.

The fact that all natural amino acids are of the L-series is important in peptide structure. If a model of the peptide chain is prepared, using the tetrahedral arrangement of carbon bonds, it turns out that the side-chains of successive amino acids stick out in opposite directions alternately. Under these circumstances there is just room along the chain for these atom groupings. The same thing would happen if all the amino acids were of the D-series.

If, however, some amino acids were of the L-series and some of the D-series, there would be inevitable places where an L-amino acid and a D-amino acid would be adjacent. Their side-chains would stick out on the same side of the chain and there simply would not be enough room for both.

For this reason the amino acids of a protein must be virtually all L or all D. The fact that they are all L may only be the result of chance, dating back to the dim time in pre-history when the first protein molecule was formed.

Enzyme molecules (themselves optically active and asymmetric, thanks to the amino acids that make them up) can naturally distinguish between an optically active substrate and its mirror image by the difference in shape, just as a right shoe will fit a right foot but not a left foot. By thus choosing between mirror images, reacting with one and not with another, enzymes help keep the body in bias.

In order for enzymes to distinguish so between mirror images, it has been shown that the substrate must attach to the enzyme molecule at three points, and this is another indirect piece of evidence in favor of the existence of the enzyme-substrate complex.

As a climax to all the evidence pointing to the temporary formation of an enzyme-substrate complex, the British biochemist Britton Chance was able, in 1949, to detect the actual presence of such a complex by means of a very delicate light-absorption test.

So far I have talked as though protein molecules were composed of nothing but amino acids. Many of them are. However, there are some proteins that, when broken down, yield substances other than amino acids, in addition to the usual amino acids. These latter are called "conjugate proteins" and the best-known example is hemoglobin, the compound that lends blood its red color.

In the 1860s the German biochemist Felix Hoppe-Seyler was able to show that hemoglobin contained a compound called "heme" as part of its structure. Heme was not an amino acid but a rather complex atom grouping containing an iron atom. Later, enzymes were discovered that also contained heme as an integral portion of the molecule.

The heme is referred to as a "prosthetic group" (from a Greek word meaning "an addition") and is essential to the functioning of the various proteins of which it forms a part. (Although the heme of hemoglobin and the various "heme enzymes" may be the same, the function of each protein is different and depends on the amino acid make-up of the remainder of the molecule. By analogy, although the cutting edge of various tools such as an ax, a scythe, and a chisel

may be the same in nature, the function of the tool depends upon its shape as a whole.)

Heme is firmly held to the remainder of the protein molecule and cannot be removed except by disrupting that molecule. In other cases, non-amino acid portions can be removed more easily and without serious harm to the protein. Certain enzymes, for instance, when purified by dialysis, lose their catalytic ability. This is not due to denaturation, for when the water that has washed the membranous bag is added to the enzyme solution, activity is regained. Apparently some essential portion of the enzyme has been removed and then returned without upsetting the delicate structure upon which activity depends.

This was first demonstrated in 1904 by the English chemists Arthur Harden and W. J. Young, who dialyzed an enzyme called "zymase" which they had obtained from yeast. The essential portion they had removed by dialysis they called "cozymase." Since cozymase could be removed easily and restored as easily, it must be held to the protein portion by loose ties indeed. Furthermore, the cozymase must be a relatively small molecule or it could not find its way through the membrane. Still further, it must be non-protein in character, for boiling did not affect it. When other enzymes were found similarly to lose small portions necessary for activity, those small portions were dubbed "coenzymes."

Beginning with the 1920s and continuing to the present day, the chemical nature of the various coenzymes was worked out. Most proved to contain phosphorus atoms as part of their molecules and many contained atom combinations of types not found elsewhere in the body. For instance, cozymase contains as part of its molecule five carbon atoms and a nitrogen atom bound into a ring of a type present in a compound called "pyridine." Pyridine does not itself occur in tissues, nor does the pyridine ring of atoms, except in several of these coenzymes. The same is true for certain atom groupings present in other coenzymes.

Enzymes are present in the body only in the small concentrations sufficient for catalysis, so that coenzymes, and in particular these unusual atom combinations, are also present in but small concentration.

This last fact gained significance in the light of another line of investigation that had been proceeding independently. All through history men had occasionally remarked upon

the connection between diet and certain diseases. In 1753, for instance, a Scottish physician named James Lind had suggested that citrus fruits and fresh vegetables in the diet would prevent scurvy. This disease, characterized by bleeding gums and increasing weakness, was the scourge of sailors who were sometimes forced to subsist, for months at a time, on a monotonous diet of salt pork and hardtack. The British Navy took to including limes in sailors' rations some decades later and wiped out scurvy on their vessels.

Nevertheless, it wasn't until 1896 that the Dutch physician Christiaan Eijkman, working in the East Indies, found that he could produce a disease similar to beriberi (common among the native population) in fowls by feeding them polished rice. He could then cure the disease by adding unpolished rice to the diet. It was the first deliberate induction of a "dietary deficiency disease."

Apparently the rice hulls removed in polishing rice contained a small quantity of some substance essential to health and life. Within the next decade biochemists found that other such substances existed in other foodstuffs. The substance in rice hulls proved to contain an amine group as part of its molecule and the Polish-born biochemist Casimir Funk, assuming that all such compounds did, called them "vitamines" ("life-amines"). When it turned out that not all such substances contained the amine group, the final "e" was dropped and the name became "vitamins."

Through the 1930s and thereafter, the vitamins were isolated and concentrated and their structure determined. As it turned out, vitamin after vitamin proved to have the unusual atom groupings present in the coenzymes and the conclusion to be drawn was clear.

An organism such as man cannot manufacture certain of the unusual atom groupings in the coenzymes because, apparently, he lacks the chemical machinery for the purpose. However, plant life generally can manufacture them and herbivorous animals pick them up from the plants they eat. Small quantities are therefore present in man's food and small quantities are all that are needed since, as I have said, coenzymes are present in the body in only small concentration.

Ordinarily a normal diet will supply man with what is needed of these compounds. However, when diet is artificially limited, as to those few foods that require no refrigera-

tion during the long ocean voyages of times past, or by the overrefinement of bread and grain, the diet lacks sufficient quantities of the vitamins. In their absence the corresponding coenzymes cannot be formed by the human body. When that happens the enzymes of which the coenzymes are an essential part are crippled; the cell chemistry begins to limp; the man becomes ill, grows progressively worse, and, if matters are carried far enough, dies.

Nor are these non-amino acid portions of proteins purely organic in nature. I have already said that heme contains an iron atom. This means that iron is essential to life since hemoglobin and the heme enzymes are. Moreover, since hemoglobin makes up such a large portion of the blood, iron is required in sizable amounts.

There are some enzymes, however, present in the body in far smaller concentration than is hemoglobin (which is not an enzyme), that contain elements found only in association with enzymes and therefore present in only vanishingly small amounts in the body. An enzyme called "carbonic anhydrase," for instance, contains zinc. Others contain manganese, copper, cobalt, or molybdenum. Only small quantities of these "trace minerals" need be present in the diet to fulfill human requirements, but their total absence disrupts the workings of key enzymes and leads to disease and, eventually, death.

Chapter 19

THE OUTER EDGES

Now that I have explained in some detail why the presence of enzymes in the lizard and their absence in the rock makes all the difference, it is reasonable to be curious as to the exact nature of the chemical reactions catalyzed by enzymes, and therefore characteristic of life.

As a matter of fact, long before the nature of enzymes and the details of their functioning were reasonably well

understood, chemists were already concerned with the chemical reactions of life. The nature of a reaction can, after all, be determined, even if one is forced to dismiss the manner in which it is brought about by the vague statement that "it is somehow catalyzed by something."

As I said in Chapter 14, the chemical changes that proceed as a result of the activity of living tissue are lumped together under the name of "metabolism."

Notice that I have defined metabolism as including those chemical changes that proceed as a result of the activity of living tissue and that I have carefully refrained from saying that they necessarily occur *within* living tissue. I have done this because there are certain important chemical reactions involving living tissue that do not take place within the tissue.

Abundant distinction? I am referring to digestion. When you swallow food, it moves down the gullet into the stomach and, eventually, intestines. It is radically changed in the process and what is finally eliminated through the anus is quite different from what entered through the mouth. While food remains in the alimentary canal (the name given to the entire complex tube running from mouth to anus; see Fig. 42) it is not really within the body. It is in a portion of the outside world, pinched off by mouth and anus. Various glands, including large ones such as the liver and pancreas and numerous small ones in the lining of the stomach and intestines, discharge juices into the alimentary canal. These mix with the food, and the enzymatic content of the juices catalyzes those metabolic reactions that we lump under the word "digestion."

Precisely because digestive changes do not take place within living tissue, these were the first to be studied (as I pointed out in Chapter 15). Digestive juices can be brought up from the canal by means of tubes, without doing more than making the experimental subject temporarily uncomfortable, and their action upon foodstuffs can then be studied in detail. There is no reason to suppose that the reaction in the test tube would be essentially different from the reaction in the alimentary canal, since both situations are equally outside the body itself.

A lucky advance in the study of digestion came about in 1822 when the American surgeon William Beaumont came across a Canadian voyageur named Alexis St. Martin who, as a result of an accidental gunshot wound, developed a permanent abdominal opening (a "fistula") leading into the

on Mackinac Island, Michigan

stomach. For ten years Beaumont was able to study the manner in which the stomach lining secreted its digestive

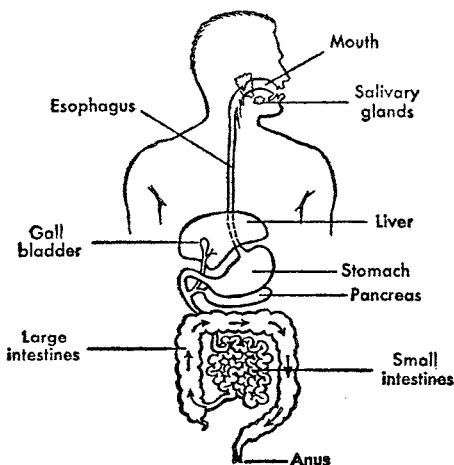


Figure 42. *The Alimentary Canal*

fluid and the effect of that fluid upon different foods under different circumstances. Beaumont published his findings in 1833 and galvanized the world of physiology.

Patients like St. Martin cannot be expected very often and to create one artificially is unthinkable. However, in the 1840s physiologists in Europe did the next best thing. They created artificial fistulas in dogs, and continued, in that way, their studies on digestion.

Through such researches, the nineteenth century saw the development of a reasonably complete knowledge concerning the chemistry of digestion, and although this was a knowledge only of what might be considered the outer edges of metabolism, it was a healthy beginning.

In summary, the knowledge gained was just this: that as foodstuffs traveled down the length of the alimentary canal one juice after another immersed it, each with its own supply of enzymes, and each taking up where previous juices had left off, assembly-line fashion. As an example of what I mean, let us consider the carbohydrates.

The chief carbohydrates in our food are cellulose, starch, sucrose, and lactose. Cellulose is indigestible in man and appears, essentially unchanged, in the feces. Starch, however, under the influence of an enzyme present in the saliva, "salivary amylase" (amylase being the general name for an enzyme catalyzing the breakdown of starch, which is *amylum* in Latin), begins to be broken down into smaller fragments called "dextrins."

Salivary amylase does not have the opportunity to do much before the food is swallowed and sent into the stomach, where the strongly acid juices put a halt to amylase activity. Another amylase, however, is found in the slightly alkaline pancreatic juice ("pancreatic amylase"), which is encountered as soon as the food leaves the stomach and enters the small intestine. Starch breakdown now proceeds in more leisurely fashion and without interruption. The breakdown to dextrins continues and the dextrins are, in turn, broken down until only small fragments consisting of two glucose units each are left. These final fragments make up "maltose."

Farther down the intestinal tract, the intestinal juices proper are encountered and these contain the enzyme "maltase," which catalyzes the breakdown of maltose to the individual glucose units.

The intestinal juice also contains "sucrase" and "lactase," which catalyze, respectively, the breakdown of sucrose to glucose and fructose, and of lactose to glucose and galactose. Thus, through the activity of co-operating enzymes, the digestible carbohydrates originally present in food are converted to the simple six-carbon units, glucose, fructose, and galactose, with glucose preponderating.

Lipids run the gantlet of fat-splitting enzymes ("lipase"), of which there is one in gastric juice, which is not very active because of the acidity of the juice, and one in pancreatic juice, which is very active. The liver secretion ("bile") contains no enzymes but does contain certain substances ("bile salts") that increase the ease with which the lipids mix with the watery digestive juices and hence the ease with which they are digested. By the time the lower portions of the small intestine are reached, lipids are broken down to glycerol (a three-carbon alcohol) and fatty acids (long hydrocarbon chains with a carboxyl group at one end).

Proteins face the most complicated ordeal of all, as one might expect. The pepsin in the gastric juice is right at

home under conditions of strong acid and is unique among enzymes in that it will not be active otherwise. It catalyzes the splitting of certain peptide bonds, while the hydrochloric acid in the juice catalyzes the breakdown of all the peptide bonds indiscriminately. The trypsin and chymotrypsin of pancreatic juice are also protein-splitting enzymes ("proteases") and take their turn when the food leaves the stomach and enters the small intestine. They catalyze the splitting of peptide bonds that are not touched by pepsin and that might have escaped the hydrochloric acid. The result is that proteins enter the main body of the small intestine in the form of a variety of small peptides made up of two, three, or four amino acids. The intestinal juice contains a variety of "peptidases" that complete the job of catalyzing the breakdown of the various peptides to single amino acids.

All these digestive changes have certain things in common. In the first place, all involve the conversion of larger molecules to smaller ones. Such metabolic changes from the relatively large and complex to the relatively small and simple are included under the term "catabolism" ("to throw downward"). Digestion, then, involves a series of catabolic changes.

Furthermore, the nature of the various catabolic changes in digestion invariably includes the splitting of some chemical bond and the attachment, at the point of splitting, of the elements of water. This method of breaking a bond by the addition of the elements of water is a natural device considering that water is the major constituent of living tissue, making up some 60 per cent of the weight and 98 per cent of the individual molecules of a human being, for instance. The process is termed "hydrolysis" (from Greek words meaning "to loosen with water") and the details of the hydrolytic change involved in the digestion of each of the three classes of foodstuffs is given in Fig. 43.

The end-products of digestion—the structural units obtained through the enzyme-catalyzed hydrolyses of the complex substances originally present in the food—can cross the intestinal membranes (be "absorbed") and enter the body itself. Molecules more complicated than the final end-products do not, in general, do so.

That is the significance of digestion: that it converts the unabsorbable foodstuffs into the absorbable structural units. What's more, it is particularly important that the original

foodstuff in all its complexity *not* be absorbed. Each organism has its own variety of carbohydrate and lipid, differing in small detail from those of other organisms. In particular each organism has its own variety of proteins. To attempt to incorporate the alien molecules of what is eaten into the structure of the eater, in totality and without modification, would seem to be impossible. In fact, when small quantities of alien proteins (or certain complex carbohydrates) are, for some reason, taken intact into the body, the organism develops special protein molecules ("antibodies") that combine with the intruders and remove them from the scene.

Such antibodies remain on the scene for further use and can serve a good purpose in rendering a person immune to further attacks of disease, where the foreign protein involved is part of an invading microorganism or part of a toxin produced by such a microorganism. Such antibodies may also serve a most inconvenient purpose in "sensitizing" an individual to the presence of alien proteins of a relatively harmless nature and thus producing allergic reactions such as those of hay fever, asthma, or food sensitivity.

When it is the structural units only with which the body is dealing, such difficulties disappear. No matter how different the proteins of cattle, fowl, or wheat may be among themselves, and how different all may be from human protein, the amino acid constituents of all are the same. As a result of digestion, the body absorbs not the foreign protein, but only the amino acid units. Once these are inside the body they can be built up into human protein. In the same way, foreign carbohydrate is broken down to glucose and built up to human carbohydrate, while foreign lipid is broken down to glycerol and fatty acids and built up to human lipid.

Once the structural units cross the intestinal membrane and are absorbed, they are within the body proper and were lost to the sight of the nineteenth-century biochemists and physiologists. The metabolic changes within the body itself ("intermediary metabolism") could be deduced only indirectly.

To be sure, the nature of certain over-all changes was clear. For instance, glucose and fatty acids must, at least in part, undergo further catabolic changes to the still simpler molecules of carbon dioxide and water. The amino acids must be catabolized, at least in part, to carbon dioxide and

water and to the nitrogen-containing urea, as well. This had to be so since carbon dioxide appeared in exhaled breath

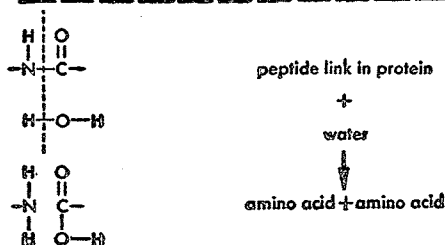
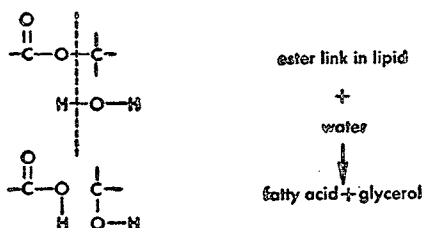
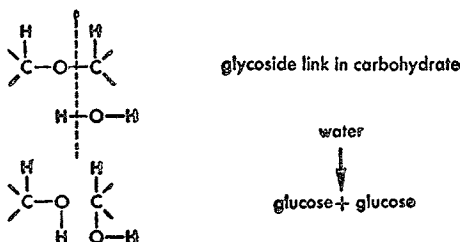


Figure 43. *Hydrolysis during Digestion*

and urea in the urine, while water was excreted by the body by way of both breath and urine and through the skin, in the form of perspiration, as well (see Fig. 44).

In addition to that, the various structural units had to, at least in part, undergo changes that would build them up into complex molecules again. This was clearly true for the amino acids and the fatty acids, since the human body con-

tained distinctly human protein and lipid that could be formed only out of the structural units.

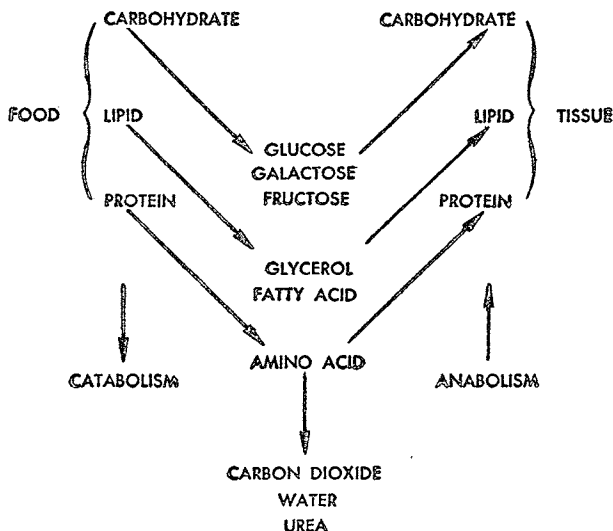


Figure 44. *Over-all Scheme of Metabolism*

The presence of distinctive human carbohydrate is less obvious. In 1844, to be sure, the German biochemist C. Schmidt found that blood contained small quantities of glucose, but this was only the structural unit itself. In 1856, however, the French physiologist Claude Bernard capped years of research by showing that the liver contained stores of a carbohydrate resembling starch, yet with certain points of difference. Bernard called this carbohydrate "glycogen" from Greek words meaning "to produce sugar" because, through hydrolysis, it could be made to do just that.

The metabolic reactions in the body that lead to the production of relatively large complex molecules from relatively small simple ones—as of protein from amino acids, lipid from fatty acid, glycogen from glucose—are examples of "anabolism" ("to throw upward").

That both anabolic and catabolic reactions are involved

in intermediary metabolism is clear and inescapable. With this much understood, then, even with the chemical details missing, much could be learned.

For instance, the glucose concentration of the blood remains virtually steady at all times. Each cubic centimeter of blood contains roughly one milligram of glucose, which serves as the immediate food supply of the cells bathed by the blood. The level does not rise unduly when the digestion of a large meal floods the body with new glucose. Nor does it fall unduly when a period of fasting temporarily deprives the body of an outside glucose supply.

This is the result of a well-organized balance of anabolism and catabolism. The glucose arising from the digestion of carbohydrates (catabolism) passes through the intestinal wall and into blood vessels that come together to form the "portal vein." The portal vein carries its heavy load of glucose to the liver, where the vein breaks up into a network of "sinusoids" that are lined with liver cells. The glucose passes from the blood into the liver cells and there it is converted to glycogen (anabolism) and stored.

The blood emerges at the other end of the liver with only a normal concentration of glucose. The hepatic vein leads that blood from the liver to the heart, from where it is pumped through the body generally. The glucose in the blood is absorbed into the various body cells and there broken down to carbon dioxide and water (catabolism).

The drain on blood glucose brought about by its catabolism within the cells is made good by the constant flow of blood from the liver with a fresh supply.

Once digestion is done, however, the flood of glucose through the intestinal wall ceases and, if fasting is prolonged, a new supply may not arrive for hours or even days. What then? When, under such conditions, the portal vein enters the sinusoids of the liver with its blood supply denuded of glucose, matters switch into reverse. There is no glucose to be stored as glycogen and therefore the glycogen already in the liver, where it had been stored during the meal of some hours past, now breaks down (catabolism), bit by bit, in just sufficient quantity to supply the blood in the hepatic vein with the proper amount of glucose.

By thus balancing catabolism and anabolism, the liver converts the alternate flood and famine of glucose into a steady and even flow. (The effect is that of a bank account in a

well-run family, in which money is stored during prosperity and from which it is removed in adversity, thus keeping life at its even tenor.)

The glucose balance in the body is a most delicate one and is maintained by the hormones "insulin" and "glucagon," both of which are produced by certain cells in the pancreas and then secreted directly into the blood. (The pancreas is thus a double gland: a digestive gland, secreting juices through a duct into the small intestine; and an endocrine gland, secreting hormones directly into the blood.)

The presence of insulin in the blood tends to encourage the anabolic reaction; that is, the conversion of glucose to glycogen; in this way lowering the concentration of glucose in the blood. Glucagon encourages the catabolic reaction; the conversion of glycogen to glucose; and raises the concentration of glucose.

If, for any reason, the body is running a little too heavily in the direction of catabolism, so that the liver is releasing a bit too much glucose into the blood, this high-glucose level, as it passes among the cells of the pancreas, stimulates the production of additional insulin. The glucose level at once drops. If it drops too far, the low-glucose level stimulates the production of glucagon, which raises the level once more. Between the two, the glucose level is held steady by means of what, in these modern days of automation, we have come to call "feedback."

Occasionally the capacity of a human pancreas to form insulin declines. The result is a disease called "diabetes mellitus." The unbalanced action of glucagon has a tendency to drive the glucose level in blood to higher values than normal; values high enough, even, to become dangerous. As the situation grows severe, the glucose level can only be reduced by removal through the kidneys. The appearance of glucose in the urine is a sure sign of well-advanced diabetes.

Before the twentieth century, there was no way of treating the disease. The physicians and biologists of the nineteenth century, to be sure, had learned to deal with many infectious diseases by means of vaccines and antitoxins, but before diabetes, they were helpless. It was not caused by any germ or by any poison. It was a failure of function of body chemistry; a disorder of metabolism; a "metabolic disease."

The proof of this came in 1889 when two German physiologists, J. von Mering and O. Minkowski, found they could produce diabetes in a dog every time by just cutting out its pancreas. The search was on for something in the pancreas that could reverse the effect when injected into the blood stream, but for more than twenty years all efforts failed.

In 1921 the Canadian physician Frederick Grant Banting guessed that this might be owing to the fact that the substance involved was a protein and that during the process of attempted extraction the proteases of the pancreas digested it. With the help of a young student named Charles Herbert Best, Banting tried tying off the duct of a pancreas before attempting to remove it and extract the desired substance. This put the digestive juices out of action and those portions of the gland that produce them atrophied. The remainder of the gland could now be extracted to yield what was eventually named insulin.

By use of insulin, prepared from the pancreases of slaughtered cattle and swine, diabetes can be controlled (but not cured, for there is no known way of causing a cell to regain a function it has lost—in this case the production of insulin) and human diabetics can lead reasonably normal lives.

Let's consider the glucose balance further. The liver can hold only so much glycogen. Once 10 to 15 per cent of its weight is glycogen, it is gorged and can hold no more. In addition, glycogen can be stored in muscle to the extent of 1 per cent. There is much more muscle in the body than there is liver, so that the total amount of glycogen in the muscles is greater than the total in the liver, despite the lower concentration in the former. All told, the liver may contain about 250 grams of glycogen and the muscles about 350 grams. A total of 600 grams is therefore all the body will hold.

Now remember that, in Chapter 14, I said the caloric value of carbohydrates comes to about 4 kilocalories per gram when converted to carbon dioxide and water. The total body glycogen content of 600 grams can therefore yield 2400 kilocalories, if totally consumed. This will last an individual one day, if he spends it quietly, and yet it is possible to fast for several weeks without dying and during all that

time blood glucose is maintained reasonably constant. Where does the glucose come from if the glycogen is used up during the first day of a fast?

An allied question arises from the fact that it is possible to eat more starch than is necessary for the energy requirements of the body and keep that up indefinitely. How is all the glucose that arises from that starch stored if the body will hold only so much?

The answer lies in lipid. We all know that potatoes, and starchy food in general, are fattening, and what does that mean but that the body can convert carbohydrate into lipid? Glucose can, apparently, be changed to some still simpler substance that can be converted to fatty acid, which fatty acid can then be built up into lipid. This is an economical thing for the body to do since a gram of fat yields 9 kilocalories, rather than the 4 of carbohydrates. Weight for weight, two and a quarter times as much energy can be stored as fat than as carbohydrate. People who overeat and swell to ungainly proportions would seem even more monstrous if, like potatoes or rice, they stored their chemical energy in the form of the energetically less compact starch.

Moreover, the change is reversible. During a fast, an individual loses his fat supply as we all know. Since the fat supply, unlike the carbohydrate supply, can be built up in virtually unlimited quantities, a fast can continue for weeks.

Under extreme conditions of fasting, when even the fat stores are scraping bottom, proteins are judiciously thrown into the fire. The amino acids, in other words, are stripped of their nitrogen content and what remains can be converted into glucose. As a result, the tissues themselves waste away during a prolonged fast.

All this gives the impression that the structural units of the various foodstuffs merge at some point in their further catabolism. Glucose and fatty acids must be capable of catabolic change to some common building block simpler than either. Furthermore, amino acids (minus their nitrogen content) must also be capable of catabolic change to this same common building block, whatever it is (see Fig. 45).

This gives the impression of complete versatility on the part of the body. From the common building block it can strike out anabolically in all directions, it would seem.

However, there are some limitations that must be remembered. The body is not capable of changing one element

to another, so the common building block contains carbon, hydrogen, and oxygen only, for in its formation from protein

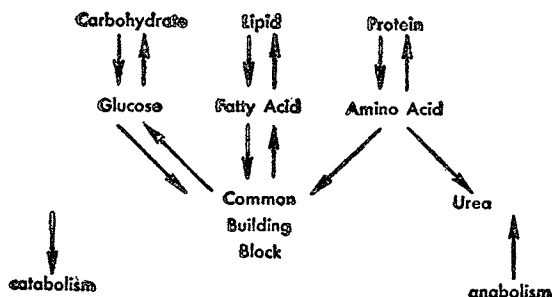


Figure 45. *The Common Building Block*

the nitrogen content of that foodstuff is removed. If the common building block does not possess nitrogen, then from it, as indicated in the diagram, no amino acid can be formed, unless, of course, there is some independent source of nitrogen. The only practical source of nitrogen in the ordinary course of events is food protein.

It is possible therefore to live on a diet that contains very little carbohydrate and lipid. This is so because, if the protein supply is ample, some amino acids can be anabolized into body protein, while the remainder can be stripped of nitrogen, converted to the common building block, and built up to carbohydrate and lipid.

The reverse is not true. A diet containing little or no protein leads to tissue wastage and death, regardless of how rich the diet is in carbohydrate and lipid. In the absence of some source of nitrogen, there is no way of converting the common building block into the necessary protein.

Even if we allow a nitrogen source in the food, as well as sources of other elements needed by the body, the versatility of the body is still limited. It cannot, as I said in the previous chapter, manufacture certain vitamins for itself. These, however, are needed in only tiny quantities and it might be tempting to dismiss this as a forgivable lack of chemical versatility. However, a similar lack on a larger scale turns up in the case of proteins themselves.

Early in the nineteenth century attempts were made in

France to prepare a cheap food by boiling bones and connective tissue (ordinarily indigestible) to produce gelatin, which is eminently digestible. The only trouble was that gelatin, although a protein, could not support life when it was the only protein in the diet.

As the nineteenth century progressed, however, chemists began to learn that proteins were composed of a variety of amino acids, and in 1872 the German biochemist O. Nasse was able to show that one of them, tyrosine, was lacking in gelatin. In later years it was shown that other amino acids such as cystine, tryptophan, and methionine were also lacking. But after all, proteins are digested to the amino acid stage and it is the amino acids and not the proteins that are absorbed. It is reasonable to suppose, then, that it is not the nutritive qualities of this or that protein that should be considered, but the nutritive qualities of this or that amino acid. The value of the protein could then be interpreted in terms of the amino acids it possessed or lacked.

Dietary experiments in which the source of nitrogen was a mixture of amino acids, rather than protein, clearly showed that life and growth could continue in normal fashion if all the amino acids were present. In fact, when the American biochemist William C. Rose first conducted such experiments and found that trouble arose, he suspected the existence of a hitherto unknown amino acid that had not been included in his mixture. In 1935 his search uncovered threonine, the last of the important amino acids to be discovered. With threonine added to the amino acid mixture, all was well (which alone is rather strong evidence in favor of the thought that no more major amino acids remain to be discovered).

Rose conducted his dietary experiment with human adults: graduate students for whom the somewhat less than exiting diet of starch, corn oil, amino acid mixture, and vitamin-mineral fortifications was rendered palatable by the thought of service to science and, no doubt, the necessity of passing a course. In these experiments Rose used as his criterion for dietary sufficiency something called "nitrogen balance."

The nitrogen content of food can be balanced by the nitrogen content of feces (food nitrogen that was never absorbed) and by the nitrogen content of the urine (food nitrogen that was absorbed but which was later stripped from body protein for one reason or another and excreted). Under ordinary conditions, in the healthy adult, the nitrogen

content of the food is just equal to the nitrogen content of the feces plus the urine. The nitrogen content of the body is then constant, since intake equals outgo, and the body is in nitrogen balance.

There is always some nitrogen outgo, however. Even when the body's dietary supply of calories is ample, so that no protein need be sacrificed for energy, there is still a certain rate at which protein is stripped of its nitrogen. (This is sometimes referred to as "wear and tear.") If this trickle of protein wastage is not replaced, nitrogen outgo is greater than intake and the body's content of nitrogen decreases. It is then in "negative nitrogen balance."

It is also possible for intake to be greater than outgo, so that the nitrogen content of the body increases and the body is said to be in "positive nitrogen balance." Growing children are invariably in positive nitrogen balance, as are people recovering from semi-starvation or a wasting disease who are building up their tissues once more after a prolonged period of negative nitrogen balance.

In order for a body to be in nitrogen balance, protein anabolism must balance protein catabolism. Protein must be synthesized from amino acids as fast as protein is used up through wear and tear. But for protein to be synthesized from amino acids, all the amino acids composing that protein (which usually means all nineteen important amino acids) must be present in the quantities needed. Nutritional experiments have shown that the absence of a single necessary amino acid makes the synthesis of the protein molecule impossible. Apparently the body cannot leave a hole in the amino acid chain to be filled later.

If a single necessary amino acid is absent, then, the protein lost through wear and tear is not replaced, and such amino acids as are present are good for nothing but energy. They are stripped of nitrogen and catabolized. The body goes into negative nitrogen balance.

But if the body must have every amino acid present, does that mean that each one must be present *in the diet*? No, for some amino acids could be omitted from the mixture fed to graduate students by Rose without their going into negative nitrogen balance. Alanine, for instance, can be left out with impunity. But alanine is necessary in protein structure and must be present in the body whether or not it is present in the diet. The conclusion is that alanine is synthe-

sized in the body, using the nitrogen supplied by the remaining amino acids.

Table 8
THE ESSENTIAL AMINO ACIDS

Amino Acid	Daily Requirements (grams)
Phenylalanine	1.1
Methionine	1.1
Leucine	1.1
Lysine	0.8
Valine	0.8
Isoleucine	0.7
Threonine	0.5
Tryptophan	0.25

A number of amino acids can thus be omitted and their synthesis in the body will take place in ample quantity. These are the "non-essential amino acids" (not non-essential for the body, mind you, but only non-essential as a component of the diet).

On the other hand, the amino acid, lysine, if omitted from the mixture, throws the student into negative nitrogen balance. The body must have lysine in its diet; it apparently cannot manufacture lysine for itself no matter how ample the supply of the remaining amino acids. Lysine is an example of an "essential amino acid" (essential as a component of the diet).

Rose found that all together eight amino acids were essential for the adult human, and measured the minimum daily quantity that could safely be assumed to meet body needs and prevent negative nitrogen balance and tissue wastage. The amino acids and the amount required are given in Table 8.

Here, then, is a larger result of the lack of complete versatility on the part of body chemistry than is the case with the vitamins. Vitamins in milligram quantities will satisfy the daily body needs, but the essential amino acids are needed in hundreds, even thousands, of times that amount.

Chapter 20

LIFE WITHOUT AIR

As you see, a number of things can be discovered about what must go on within tissues from a mere consideration of that which enters the body and that which leaves. We have deduced that there must be a common building block through which carbohydrates, lipids, and proteins must all pass. We can say that the structural units of foodstuffs—glucose, fatty acids, amino acids, and so on—can be either catabolized to carbon dioxide and water or anabolized to starch, lipid, or protein. We can even decide that specific substances, such as alanine, can be synthesized within the body while others, such as lysine, cannot.

Have we exhausted all our resources now or can we deduce still more?

Well, we might reason as follows: An enzyme can only catalyze a reaction to the point of equilibrium, and it will do so beginning at either end of the reaction (see Chapter 12). Therefore, when digestive enzymes break down complex foodstuffs by hydrolysis, they surely do so only to the equilibrium point of the reaction. Yet it is also true that hydrolysis is complete, that virtually all the complex components of food are hydrolyzed and absorbed.

But that is the point; they are absorbed. As the amino acids, fatty acids, and simple sugars formed by hydrolysis are absorbed, they are removed from the scene of action. Only the complex compounds remain behind. These are further broken down and so on. In other words, the law of mass action (see Chapter 9) acts, through removal of the end products, to shift the equilibrium far in the direction of hydrolysis.

Can we not use this to deduce what must go on in cells? Within the tissues, only the hydrolyzed end products present, we could picture the same enzymes as catalyzing the reverse

reaction to the same equilibrium point as before. Instead of bonds being broken with the addition of the elements of water, a hydrogen is taken from one amino acid (let us say) and an oxygen-hydrogen combination from another, the two bonds so exposed uniting to form a peptide link (see equation on p. 233). Such an opposite-of-hydrolysis reaction is called a "condensation." If the cell could now remove the products of reaction (in this case, protein and water) as quickly as they are formed, condensation could proceed as far as necessary, even to completion.

This is a pretty picture, with the same enzymes controlling both catabolism and anabolism according to which sets of products are removed from the scene of action—but is it true?

One point in its favor is that digestive enzymes can be made to catalyze anabolic reactions in the test tube. For instance, if amino acids are left exposed under proper conditions to the action of pepsin or chymotrypsin, some condensation does indeed take place and an insoluble protein-like substance called "plastein" is formed. Plastein is a mixture of rather low-molecular-weight peptides, with chains consisting of perhaps no more than ten or twelve amino acids, but it is permissible to suppose that conditions within tissues would be more efficiently adjusted to protein formation than are conditions within a test tube and that larger and more complex proteins could then be formed.

But, alas, there are serious catches to this whole theory. In the first place, digestive enzymes of the type found in the alimentary canal simply do not exist within the tissues. Whether pepsin or chymotrypsin can form protein out of amino acids is beside the point, if neither enzyme (or any enzyme similar) is to be found in the cells where protein synthesis takes place.

Again, there is the matter of free energy. On the whole, catabolic reactions are accompanied by decreases in free energy. In the case of hydrolyses, the decrease is not a large one (at least as compared with some other types of catabolic reactions I shall mention later) but it is large enough. In the case of the hydrolysis of the peptide bond, the decrease in free energy varies from 0.5 to 4 kilocalories per mole, the variation depending on the nature of the amino acids involved in the bond.

This means that the hydrolysis of proteins is a downhill

reaction with an equilibrium point far over toward complete hydrolysis. The free-energy decrease is large enough so that the reaction is not markedly reversible. In fact, at equilibrium point, roughly 99 per cent of the protein is hydrolyzed.

This doesn't mean that the reaction cannot be reversed under appropriate conditions. In the test tube, conditions are arranged so that water is removed as fast as it is formed (water being one of the products of a condensation). The reaction is driven in the direction of protein, and plastein is formed. In the cell, however, water—present in such overwhelming preponderance—cannot be removed, and the removal of protein from the scene of action as fast as formed presents difficulties as well. No doubt if this were the only possible pathway, we might expect to find some way in which the body accomplishes the task, for living tissue accomplishes many surprising things. However, should we not ask ourselves first if there is any reasonable alternative to a mere reversal of reaction?

Well, if catabolic reactions are generally accompanied by decreases in free energy, then the reverse anabolic reactions are generally accompanied by increases in free energy. Anabolism is an uphill process. Rather than deal with the tiny bit of downhill reaction at the other side of the equilibrium point, working madly to remove products in order to keep it going, why not abandon the equilibrium altogether and simply push the reaction uphill by the direct application of energy? In that case the reaction could go in either direction according to whether or not energy was applied—in the direction of catabolism in the absence of applied energy; in the direction of anabolism in the presence of applied energy.

Where is the energy to come from, however? The obvious source lies in the further catabolism of the structural units. The energy produced by the conversion of glucose to carbon dioxide and water is, weight for weight, hundreds of times greater than the energy produced by the hydrolysis of glycogen to glucose (or than the energy required for the condensation of glucose to glycogen). Therefore, by sacrificing one glucose molecule to catabolism, it is conceivable that hundreds of other glucose molecules might be forced uphill to glycogen.

But just how and in what manner is the energy of catabolism trapped and made use of in driving the energy-consuming anabolic reactions? To answer that question, catab-

olism must be studied in detail. There is no escaping it any longer. We must find out what happens to foodstuffs after their structural units disappear into the cell. We must penetrate the mysteries of intermediary metabolism.

One form of catabolism that interested scientists quite early and that served as a starting point in the investigation of intermediary metabolism was that in which yeast cells fermented glucose. For one thing, yeast fermentation of sugars has been important to mankind since prehistoric days, so that there were practical reasons for advancing knowledge concerning it. For another, yeast fermentation is a relatively simple system that can be studied without the complications involved in dealing with the tissues of animals.

The fermenting action of yeast serves to break down glucose to ethyl alcohol, liberating carbon dioxide, so that the equation can be written this way:



Notice that this equation balances. Each molecule of glucose is catabolized to two molecules of ethyl alcohol and two molecules of carbon dioxide. Nothing else is involved.

In a way, this is surprising. After Lavoisier had demonstrated that animal life could not exist without oxygen, it was naturally felt that important catabolic reactions must involve oxygen. This one didn't.

Well, then, surely other reactions brought about by yeast did? The negative answer to this came through the researches of Louis Pasteur, who in 1861 was able to show that yeast carried on all its life processes in the complete absence of oxygen. It was *la vie sans air*, or "life without air."

The conversion of glucose to ethyl alcohol and carbon dioxide is an example of "glycolysis" (which, in Greek, means, loosely, "the breaking up of sugar") and, to emphasize the fact that it takes place "without air," it is often called "anaerobic glycolysis," "anaerobic" meaning "without air" in Greek.

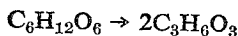
What with man's experience in burning fuel, it is natural to associate the formation of energy with the combination of carbon-containing substances with oxygen, but we must not be over-rigorous in this respect. Energy can be produced

in other ways too. The breakdown of one mole of glucose to ethyl alcohol and carbon dioxide liberates about 36 kilocalories of free energy. This is, to be sure, much less than can be liberated by the complete conversion of that same mole of glucose to water and carbon dioxide. In the latter case, the free-energy change is 686 kilocalories, so that glycolysis produces only a little over a twentieth of the energy available in the glucose. Nevertheless, glycolysis offers enough to live on—for yeast.

Furthermore, it seems quite likely that, at one time early in the history of the development of life, the atmosphere of earth was free of oxygen, so that what energy was available to organisms had to be developed by methods akin to glycolysis. Even today, complex organisms that ordinarily get their energy by catabolizing glucose all the way to carbon dioxide and water, consuming oxygen in the process, remain capable, under specialized circumstances, of getting along temporarily by way of glycolysis.

Muscle tissue is a case in point. At the beginning of the twentieth century it was discovered that muscle contained small quantities of lactic acid, a three-carbon compound with the empirical formula, $C_3H_6O_3$. (Previously, this compound had been found in sour milk and it is from the Latin word for milk that it obtains its name.) It was also discovered that when muscle went into action its content of lactic acid increased.

In 1918 the German biochemist Otto Meyerhof was able to show that lactic acid increase was accompanied by a decrease in the glycogen stored in muscle. It seemed clear that muscular activity involved the breakdown of glycogen to glucose and the conversion of the latter to lactic acid. The following equation for the conversion of glucose to lactic acid can be written:



Again, the equation is balanced. One molecule of glucose is broken down to two molecules of lactic acid and oxygen is not involved. It is this, or something very similar, that must be happening in souring milk, with infesting bacteria converting milk sugar to lactose. The results in muscle, or in souring milk, are different from the results in yeast, since

in the latter case lactic acid is subjected to a further split:



to ethyl alcohol and carbon dioxide. Energetically, however, this is a trifling detail, since no significant amount of additional energy is developed by this last "decarboxylation" (i.e., the splitting off of carbon dioxide from a compound).

The failure of muscle to perform the last step is a good thing, of course. Alcohol is more toxic than lactic acid and its presence in the cell would be correspondingly more dangerous. It is all very well for yeast cells to form alcohol, for they discharge it into the surrounding medium; and even so they are killed when the alcohol concentration rises to about 18 per cent. Muscle cells would have to discharge alcohol into the blood stream and the organism would quickly get gloriously drunk and then die. To be sure, our chemistry might have become adapted to high concentrations of alcohol, but it didn't. Apparently the method chosen by developing multicellular life, in dealing with the problem of alcohol, was to get rid of it altogether by eliminating the final step of glycolysis.

Meyerhof did more than just show that complex organisms were capable of bringing about anaerobic glycolysis, even though they also possessed the ability to make efficient use of oxygen. He also showed that yeast and muscle possessed the same coenzymes. This made it appear possible that both types of cells broke down glucose by a similar series of reactions. It was the beginning of the demonstration that the metabolic pathways of all organisms were essentially similar and that what differences did exist were merely variations on a theme. All researches since Meyerhof's time have strengthened this view.

It is easy to reason out the usefulness of this ability on the part of muscle to glycolyze glucose. The muscle can obtain its energy by the complete conversion of glucose to carbon dioxide and water and does so under conditions of peace and quiet. In order to do this, however, it naturally requires oxygen. The oxygen is brought to the muscle by the blood stream and, if complete conversion of glucose to carbon dioxide is to be carried out, the muscle's rate of

doing work would be limited by the rate at which oxygen could be brought to it.

When muscle tissue begins to expend energy at an increased rate, as when you exercise, the body speeds up oxygen transport. You breathe faster and deeper and your heart pumps both faster and harder. This naturally increases the rate at which muscle receives oxygen and, therefore, can develop energy. But there is a limit to how quickly you can breathe and how quickly your heart can beat. What if it is vitally necessary that the muscle work at a rate beyond that which can be supported by the oxygen supplied by lungs and heart even when the latter are working at peak load?

That might be stamped "impossible," but the evolutionary race goes to the organism that can do just a bit more than the next, and it would be profitable, and of survival value, if the muscle retained the primitive ability of anaerobic glycolysis. It could then obtain additional energy without requiring additional oxygen.

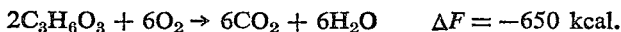
This is inefficient, of course. Some nineteen molecules of glucose must be broken down to lactic acid to produce what one molecule broken down to water and carbon dioxide would produce. But in an emergency, that prodigal waste of glucose is preferable to the lack of that last bit of drive needed to save your life, perhaps.

The process is also limited. Lactic acid accumulates as the muscle works and, although lactic acid is not as toxic as alcohol, it can be tolerated only up to certain limits. Resting muscle contains 0.02 per cent lactic acid. During work, this may rise to 0.25 per cent, but not much more. We recognize the rise of lactic acid concentration in the form of fatigue, and eventually the muscle, no matter how driven, must stop. It can take no more lactic acid.

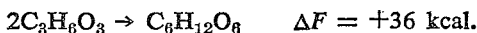
What happens to the lactic acid, then? Obviously, the muscle recovers after resting. It was the British physiologist Archibald Vivian Hill who showed that muscle, after having become charged with lactic acid, consumes oxygen while resting, at a rate greater than normal. The oxygen that would ordinarily have produced energy by combining with glucose, and which was short-circuited during the glycolysis, is now making up for what it had missed. We have incurred an "oxygen debt" and that accounts for the fact that long after

we have worked ourselves to exhaustion and collapsed in a heap we continue panting heavily while our heart continues to thud away. We are supplying the oxygen as rapidly as possible to discharge the debt.

The debt is discharged by the breakdown of lactic acid to water and carbon dioxide. This produces the remaining eighteen-nineteenths of the energy that can be produced from the complete breakdown of glucose:



The energy thus released is much greater than the energy that would be required to send two lactic acid molecules back uphill to glucose:



If the energy produced by lactic acid catabolism were harnessed and put to work converting other lactic acids to glucose, you can see that one catabolized lactic acid molecule would suffice to anabolize many lactic acid molecules. Even if the energy were used with only 50 per cent efficiency, one lactic acid molecule converted to carbon dioxide and water would supply the energy necessary to reconvert ten lactic acid molecules to glucose.

In the end, with the oxygen debt discharged, most of the lactic acid is back as glycogen, ready for another bout of work, while the amount of glycogen lost through complete catabolism in the process is easily replaced out of the endless (during life, at least) supply of blood glucose.

The downhill catabolism of lactic acid, making possible the uphill anabolism of lactic acid, is an example of a "coupled reaction" (see Fig. 46). In the over-all reaction there is a decrease in free energy. That is, although the change of lactic acid to glucose involves an increase in free energy, it takes place only in response to the change of lactic acid to carbon dioxide and water, involving a decrease in free energy. And the decrease in one portion of the reaction is always larger than the increase in the other, so that the net change is always a decrease, as it must be *even when living organisms are involved*.

It is all very well to speak of the coupling of lactic acid

catabolism with lactic acid anabolism, but exactly how are they coupled? In what manner is energy shifted from the former to the latter?

The first hint came in 1905. The English chemists Harden and Young (discoverers of zymase) found in that year that when the rate at which yeast cells produced carbon dioxide began to fall off it could be restored to its original speed by the addition of an inorganic phosphate. (An example of an inorganic phosphate is potassium acid phosphate, KH_2PO_4 .) Furthermore, as carbon dioxide evolution proceeded, the quantity of the phosphate remaining in solution decreased.

It seemed to Harden and Young that the only way in which phosphate could disappear was by its combination with some organic compound. The phosphate would then no longer exist in simple ionic form and it would no longer react with the test reagent. They began to analyze the mixture for small quantities of any organic compound that might contain phosphate groups and found one. The phosphate group was bound to a sugar molecule by a kind of link characteristic of those compounds known to chemists as "esters." Consequently it was called the "Harden-Young ester" to begin with. With time, the exact nature of the compound was determined and it turned out to be a molecule of the simple sugar, fructose, to which two phosphate groups were attached. It is now called "fructose diphosphate."

Fructose diphosphate is an example of a "metabolic intermediate," a compound that occurs in a position in the scheme of metabolism between some easily detected starting point (in this case, glucose) and some easily detected end point (in this case, ethyl alcohol and carbon dioxide). The detection and identification of metabolic intermediates is crucial to any understanding of the details of metabolism within the tissues, so it may be said that 1905 marked the birth of the systematic study of intermediary metabolism.

Other organic phosphates were discovered shortly thereafter and, to the surprise of everybody, it became more and more evident that the phosphate group was intimately connected with every detail of metabolism. Until then the phosphate group had been known to occur in bones as part of the inorganic framework of the body. Its presence in the soft tissues, where its function was not understood, had been largely ignored. Now, however, phosphate began to take

the center of the stage, and it is therefore worth a chapter to itself.

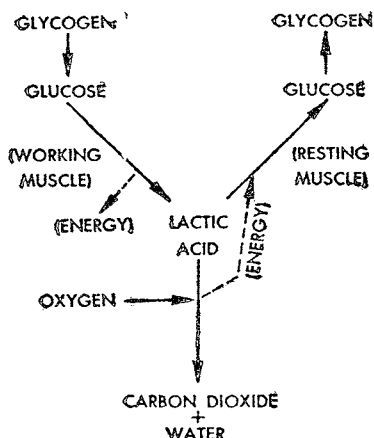


Figure 46. *Lactic Acid and Muscle*

Chapter 21

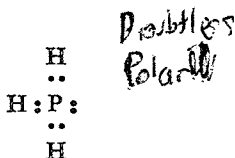
THE IMPORTANCE OF PHOSPHATE

In order to discuss the connection of the phosphate group with metabolism, we had best begin with the element, phosphorus, itself.

Phosphorus appears just below nitrogen in the periodic table, so the atomic structures of these two elements are similar. The nitrogen atom has two shells of electrons with five electrons in the outer shell, while the phosphorus atom has three shells of electrons with five electrons in the outer shell. Electronically, then, the phosphorus atom can be rep-

resented as $\cdot\overset{\cdot}{\underset{\cdot}{\text{P}}}\cdot$, a representation analogous to that for nitrogen.

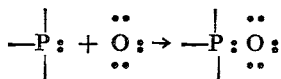
The phosphorus atom can share each of three of its electrons with other atoms, accepting a share in three of theirs, so that the stable configuration of eight is obtained. Thus, with hydrogen, phosphorus will form the compound, phosphine (PH_3)



which is analogous in structure to ammonia (see Chapter 17).

In addition to the three electrons that phosphorus shares with other atoms, it also has a pair of electrons that it does not have to share to attain the stable configuration of eight. It doesn't share them in the case of phosphine, for instance. This unshared pair of electrons can, however, be donated, so to speak, to an atom that happens to be short two electrons to make up the total of eight that it requires. The atom we think of first in this connection is the oxygen atom with only six electrons in its outer shell.

If, for the moment, we just concentrate on phosphorus and oxygen, we can let the ordinary pooled electrons of the phosphorus atom be represented by the usual valence bonds and consider only the unshared electrons, thus:



As you see, the oxygen atom now satisfies its requirement for eight electrons, without this representing any loss to the phosphorus, which retains the donated electrons in its own outermost shell as well.

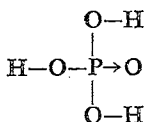
A bond formed in this manner by a pair of electrons, where both electrons are donated by a single atom, is just as strong as one formed by the usual method of having each

atom donate one of the pair. Such a one-atom-contribution bond is given the special name of "coordinate bond" and it is often represented by a little arrow, with its head pointing away from the atom that donates the electrons and toward the atom that accepts them. The valence situation for phosphorus is thus three covalent plus one coordinate and the atom can be represented as follows:



Often, as in the case of phosphine, only the three covalent bonds need be involved. However, particularly where oxygen atoms are involved, the single coordinate bond is brought into play as well.¹ (All this, by the way, is true of the nitrogen atom as well.)

The most important compound of phosphorus is "orthophosphoric acid," usually called simply "phosphoric acid" (H_3PO_4), in which all four bonds of the phosphorus atom are attached to oxygen atoms and which can be represented as follows:



Wherever the phosphorus atom occurs in the compounds within tissues, it always occurs in a grouping related to phosphoric acid. Phosphoric acid will ionize, for instance, losing first one hydrogen ion (H^+) to leave behind the "primary phosphate ion" (H_2PO_4^-) with a single negative charge, then a second and third hydrogen ion to leave "secondary phosphate ion" (HPO_4^{2-}) and "tertiary phosphate ion" (PO_4^{3-}). The first two of these ions are found in blood and tissues.

In addition, phosphoric acid, or the ions derived from it, can combine with other compounds. Involved in such combinations are one (occasionally more than one) of the oxygen-

¹ Because oxygen has a valence of two, the coordinate bond used to be considered a double bond



in the days before electronic structure was well understood. It was this that led to the belief that phosphorus (and nitrogen) displayed a valence of five.

hydrogen ("hydroxyl") groups present in the molecule. We can concentrate on that portion of the phosphoric acid molecule and, for brevity and convenience, indicate the remainder of the molecule by an encircled P. The shorthand formula for phosphoric acid would then be $\text{H}-\text{O}-\textcircled{\text{P}}$. (Where more than one hydroxyl group is involved, a second and third hydroxyl group can be shown attached to the central $\textcircled{\text{P}}$.)

In combining, the phosphoric acid, or its ions, usually combines with the hydroxyl group of another compound. Such a compound can be written as follows: $\text{R}-\text{O}-\text{H}$. In such a formula, "R" represents any carbon-containing group (it stands for "radical" for historical reasons we need not go into). Thus, $\text{R}-\text{O}-\text{H}$ can stand for an alcohol, or for a sugar, or for other compounds we haven't discussed; for any compound with a hydroxyl group in fact.

The reaction of phosphoric acid and, let us say, a sugar can therefore be represented as in Fig. 47. In combining, a molecule of water is split out, which makes the reaction a condensation, and its reverse is, of course, a hydrolysis.

As usual, it is the hydrolysis that represents the loss in free energy, so that the equilibrium of the reaction is far in the direction of the free sugar plus phosphoric acid. The free energy liberated in the hydrolysis of a sugar phosphate (of which an example is the Harden-Young ester, mentioned at the end of the last chapter) is from 2 to 4 kilocalories per mole, depending on the exact structure of the particular sugar phosphate. This means that hydrolyzing the ester link of a sugar phosphate liberates roughly the same amount of free energy as does hydrolyzing the peptide link of a protein molecule.

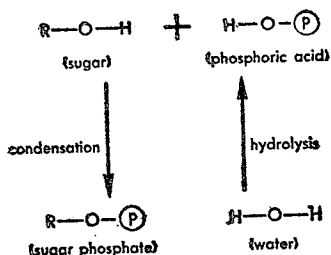


Figure 47. *Phosphoric Acid and Sugar*

Naturally to form the ester by condensing a sugar and a phosphoric acid requires the input of 2 to 4 kilocalories per mole. Therefore, forming an ester link of this type requires roughly the same quantity of energy as does forming a peptide link.

Since phosphoric acid possesses a hydroxyl group, one molecule of the acid can condense with another like itself. The resulting compounds contain the atom combination —O—(P) , a "pyrophosphate bond."

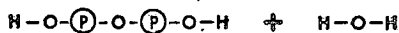
The condensed acids shown in Fig. 48 might have remained strictly in the domain of inorganic chemistry were it not for the fact that in 1929 a German biochemist, K. Lohmann, isolated a substance from muscle which, it eventually turned out, was a condensation product of triphosphoric acid with an organic compound called "adenosine" already known to occur in certain compounds called nucleic acids.² The condensation product is called "adenosine triphosphate" and is almost invariably abbreviated ATP.

Little by little, as the years passed, ATP was found to be involved in numerous types of reactions in the body. It was involved in the contraction of muscles; in the condition of impulses by nerves; in almost every branch of metabolism.

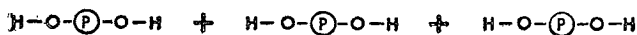
Why was this so? Well, to begin with, the two phosphate groups at the end away from the adenosine portion are more easily hydrolyzed than phosphate bonds generally are. This would indicate that the free energy loss in hydrolyzing the pyrophosphate bond might be greater than that involved in hydrolyzing an ordinary phosphate bond. Investigations proved this to be true. The German-born American biochemist Fritz Lipmann, as a result of research into the matter in the early 1940s, suggested that there were two types of phosphate bonds. The ordinary kind, such as existed in a sugar phosphate, was a "low-energy phosphate bond," while the pyrophosphate link was an example of a "high-energy phosphate bond."

² The nucleic acids are not greatly involved in the interplay of energy and life and, therefore, do not enter, except glancingly, into the scheme of this book. They are, nevertheless, the most important substances in living tissue and may even, if one is in a melodramatic mood, be equated with life itself. They control, for instance, the entropy-decreasing process of synthesis of specific proteins. If you are curious about them, you will find a discussion of these compounds in the final third of my book, *The Wellsprings of Life* (Abelard-Schuman, 1960).

At first it appeared that the high-energy phosphate bond might liberate as much as 11 to 16 kilocalories per mole



pyrophosphoric acid



triphosphoric acid

Figure 48. *The Pyrophosphate Bond*

on hydrolysis, so that, compared with the 2 to 4 kilocalories per mole liberated by ordinary phosphate bonds, there seemed a large gap between the two varieties. Later work, however, showed this to be an exaggeration. Values for the free energy loss on the hydrolysis of high-energy phosphate bonds, as obtained in the mid-1950s, proved to be no more than 5 to 10 kilocalories, with an average of about 8. Further, some supposedly low-energy phosphate bonds showed free energy changes, on hydrolysis, of up to 6 kilocalories per mole. Consequently there was no gap at all.

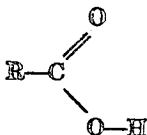
Still, we might say this. There is a type of phosphate bond that, on hydrolysis, liberates roughly 8 kilocalories per mole, and there is another type of phosphate bond that, on hydrolysis, liberates roughly 4 kilocalories per mole. Although the distinction is not as sharp as it was once thought to be, and although some biochemists are beginning to become chary of making any distinction at all, I feel it is still fair to speak of high-energy and low-energy phosphate bonds.

Many biochemists customarily distinguish between a high-energy and a low-energy phosphate bond by the manner in which they write the valence bond. A low-energy phosphate bond is written in the usual way, $\text{R}-\text{O}-\textcircled{\text{P}}$, while a

high-energy phosphate bond is written with a wiggle. Thus, if we let the adenosine portion of ATP be represented by the letter A, its formula could be written: $A-O-\textcircled{P}-O\sim\textcircled{P}-O\sim\textcircled{P}$.

The last two bonds, you see, are pyrophosphate bonds and high-energy. If one is hydrolyzed away, what is left is "adenosine diphosphate" or ADP, which is $A-O-\textcircled{P}-O\sim\textcircled{P}$. If another is hydrolyzed away, there remains "adenosine monophosphate" or AMP (which is also called "adenylic acid") and which is $A-O-\textcircled{P}$. There remains now only a low-energy phosphate bond. It can be hydrolyzed also, leaving simply adenosine, but this third hydrolysis yields only half the energy of either of the first two.

There are varieties of high-energy phosphate bonds other than the pyrophosphate. Consider an organic acid (which can be written with the general formula



where R is any of countless carbon-containing groupings and the COOH group is the "carboxyl group" which we have already met with a number of times). If phosphoric acid is condensed with the hydroxyl portion of the carboxyl group (as shown in the accompanying diagram) the result is an "acyl phosphate," written with the wiggle to indicate it is high-energy (see Fig. 49).

Now let's see how the high-energy bond comes into play, and for that, we shall return to glycogen. The glycogen in animal liver and muscle is not quite like the starch found in plant storage organs in fine structural detail, but both glycogen and plant starch are formed of glucose molecules that have been condensed, with elimination of water. When starch is digested in the alimentary canal, the amylases hydrolyze those glucose links, with addition of water again, producing smaller and smaller chains and, eventually, ending with glucose itself.

Now in the liver and muscle it is necessary to convert glycogen to glucose, on occasion, to keep the glucose level

in blood up to the mark. It would seem natural to suppose that some enzyme like the amylases would be at work in tissues, too, catalyzing the hydrolysis.

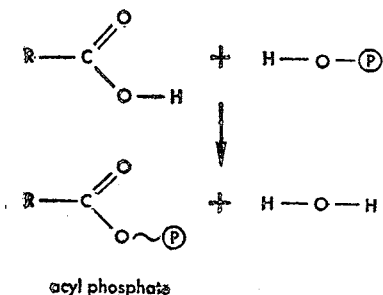


Figure 49. *Acyl Phosphate*

To check whether or not this is so, some way must be found to experiment with the reactions going on in tissues. The most direct way is to use some convenient animal, such as a mouse, as a kind of living test tube. Unfortunately, even a mouse is dreadfully complicated and thousands of reactions go on within it constantly.

The next step is to simplify the mouse, so to speak, by killing it and using small slivers of its liver or muscle. These scraps of tissue still contain working enzymes (if they are obtained properly and used quickly) and they can be soaked in solutions ("incubated") containing various chemicals. Enzymes are so specific that if a particular tissue catalyzes a certain reaction in a test tube you can be reasonably certain that it would catalyze that reaction, or one very like it, in the intact organism.

If it is necessary to speed things up, tissue can be mashed up into a semi-liquid "homogenate," which may then be incubated with substrate. There is then no necessity to wait for substrate to diffuse into cells; all the enzymes have been stripped of membrane protection by the process of homogenation and their catalytic properties go into action quickly.

Homogenates contain large numbers of different enzymes and it is usually desired to cut these numbers down; in order that there may be fewer confusing side-reactions, and in order that the desired enzyme may be obtained in greater

concentration than it exists naturally in the body so that the reaction it catalyzes may be studied more easily.

This is done by subjecting the homogenate to conditions that are known to separate one enzyme from another. Soaking it in certain solutions may bring some proteins into solution and leave others behind. Filtering the mash will then produce a solution containing only the soluble enzymes, a "tissue extract." The extract can be subjected to changes in acidity or in the concentration of certain ions or to certain organic solvents at low temperatures. Any of these conditions will precipitate some enzymes and leave others in solution. With sufficient perseverance and luck, it may even be possible to obtain a solution of a single enzyme in a relatively pure state.

It was by working with tissue extracts, for instance, that a Czech-born American man-and-wife team of biochemists, Carl F. Cori and Gerty Cori, managed to show that the breakdown of glycogen in tissues was *not* a matter of hydrolysis and was *not* catalyzed by amylases.

They found, in 1935, that if they incubated muscle extract with glycogen there was no breakdown unless phosphate ion were added. Naturally, it seemed to them that the product of breakdown would have to incorporate the phosphate ion, and sure enough, they eventually isolated a sugar phosphate (but *not* glucose) from the mixture. In 1937 they showed that this sugar phosphate, originally termed "Cori ester," was "glucose-1-phosphate." (The six carbon atoms of the glucose molecule are numbered from 1 to 6 by an internationally accepted convention and glucose-1-phosphate is a molecule in which the phosphate group is attached to carbon-1.)

Thus, glycogen in the body is not hydrolyzed, but is instead broken down by the use of phosphoric acid. I can show what I mean in the accompanying diagram. Let G stand for a glucose unit, so that glycogen becomes a string of Gs connected by oxygen links.

The long glucose chains of glycogen are broken up into single sugar units by the addition, across the links, of the elements of phosphoric acid. Now if it had been the elements of water that had been added, the reaction would have been a hydrolysis. Naturally, then, since phosphoric acid is involved instead, it is a "phosphorolysis" (see Fig. 50). Glyco-

gen is thus phosphorylated to glucose-1-phosphate in the tissues.

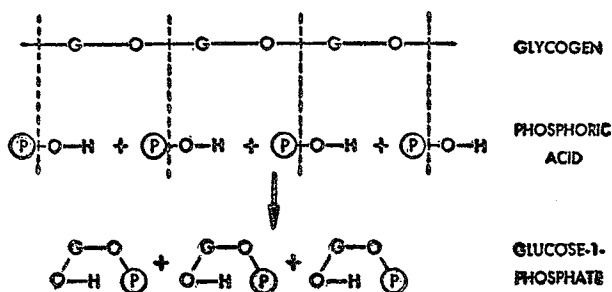


Figure 50. *The Phosphorolysis of Glycogen*

The difference in the two processes is a matter of free energy. The hydrolysis of glycogen to glucose involves a fair loss of free energy. Glycogen slides downhill and to get glucose back to glycogen by direct condensation would require the input of that amount of free energy for each link formed.

The phosphorolysis of glycogen to glucose-1-phosphate involves a very small change in free energy. Glycogen and glucose-1-phosphate are on virtually the same free energy level. The reaction is therefore easily reversible, so that by adjusting concentrations the organism can easily swing the reaction toward the breakdown of glycogen or toward its buildup, according to need, thus maintaining the balance described in Chapter 19, which is so essential to health.

And yet the production of glucose-1-phosphate does not solve all problems. For one thing, sugar phosphates cannot, for some reason, cross the barrier of the cell membrane. Any compound such as glucose-1-phosphate, formed within the liver cell, must stay within that cell until and unless the phosphate is stripped off. It is for that reason that glucose itself must be formed in digestion, for only glucose could cross the intestinal membranes and enter the body cells; glucose-1-phosphate would not. Again only glucose can be present in the blood stream, since only glucose, and not glucose-1-phosphate, can diffuse into the body cells from blood and serve as food. It is therefore glucose and not

glucose-1-phosphate that the liver cells must deliver to the blood.

This can be done. An enzyme was found in liver cells that catalyzes the conversion of glucose-1-phosphate to glucose-6-phosphate (the phosphate group being shifted from carbon-1 to carbon-6, a process requiring no energy to speak of) and then a second enzyme was found that hydrolyzed the phosphate group away, liberating glucose itself, which could then pour across the cell membrane and into the blood stream.

But you can't get around the requirements of free energy. In going from glucose-6-phosphate to glucose, you have gone down the free energy hill. Moving from glycogen to glucose involves the same loss of free energy, after all, whether you make the change directly or do it via a couple of sugar phosphates. Are you not then in just as bad a way as before? The body must find energy with which to form glucose phosphates out of glucose before it can go ahead to form glycogen. Might it not just as well find the energy to form the glycogen directly?

Apparently not. In forming glycogen, the body is faced with two major problems. It must form a large, complicated molecule out of a simple structural unit. It must also find energy for each linking of the structural unit and there are thousands of those in each molecule of glycogen. If the body were to convert glucose directly to glycogen it would have to solve both problems simultaneously and apparently it cannot. Or at least it has not, for in no sample of living tissue is glucose ever converted directly into glycogen.

If, on the other hand, glucose is first converted to glucose-6-phosphate, the energy must be found, true; but a simple compound is converted into another compound almost as simple. The problem is met and solved in isolation, without the added complication of having to form a giant molecule. Then, once the glucose phosphate is formed, it can be converted into a giant molecule without any need to find special sources of energy. The double problem is converted into two single problems, and this the body can handle.

The glycogen-glucose balance as described so far is shown schematically in Fig. 51.

The next problem is this: Since at one point or another, the body must move up the energy hill in passing from glu-

cose to glycogen, where does it find the energy that will enable it to do so? In other words, referring to Fig. 51,

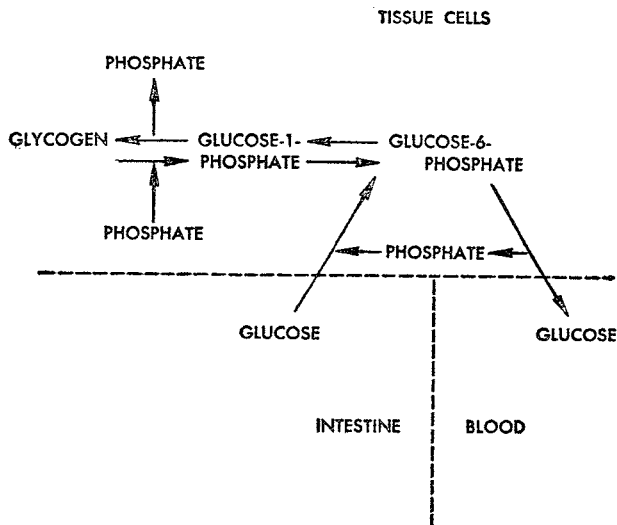
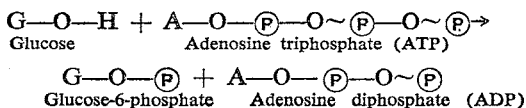


Figure 51. *The Glycogen-Glucose Bond*

how does the body convert glucose to glucose-6-phosphate?

Here is where the high-energy phosphate bond comes in. In passing from glucose to glucose-6-phosphate, two things are required, a phosphate group and an energy supply, and ATP possesses *both*.

When glucose reacts with ATP the following equation represents the results.



Consider this equation as representing a combination of two reactions. In the first, glucose is converted to glucose-6-phosphate. A low-energy phosphate bond has been formed which requires the input per mole of about 4 kilocalories.

In the second, ATP has been converted to ADP by way of the hydrolysis of a high-energy phosphate bond, a process that releases, per mole, about 8 kilocalories. The downhill momentum, so to speak, of the hydrolysis of ATP is more than enough to shove the glucose uphill to glucose-6-phosphate.

This is an example of a specific coupled reaction, with the loss of free energy in one part greater than the gain in free energy in the other, so that the net change is a loss, as it always must be. (You simply cannot get around the second law of thermodynamics.)

Until very recently there was a suspicion that ATP was *the* energy concentrate of the body; that whenever an uphill reaction was required, ATP was promptly hitched to it. Modern methods of analyzing tissue have revealed other triphosphates, however, and now it may be assumed that different reactions may require particular energy sources. The principle remains the same, however, and what I have said for ATP goes for its cousins as well.

Fructose and galactose, the other two simple sugars absorbed by the body, are also kicked uphill through the action of ATP (and by the action of specific enzymes, of course, that differ from those involved in the handling of glucose). The fructose-1-phosphate and galactose-1-phosphate are converted to glucose-1-phosphate by reactions that require no energy to speak of and in this way fructose and galactose, too, can be changed to glycogen (and, by reversal of the reactions, formed from glycogen).

The formation of glycogen from simple sugars in this fashion is called "glycogenesis." The breakdown of glycogen to the simple sugars by the same route is called "glycogenolysis." Fig. 52 presents the scheme, with ATP and all three sugars included.

But we are getting along too well. Surely we cannot simply assume that the energy supply of ATP, and molecules like it, is the final answer to everything. The body contains only a vanishingly small quantity of ATP. What happens when it is used up?

Clearly, it cannot be used up because, if it were, anabolism would cease at once and life would be impossible; as impossible for the lizard as for the rock. Since life continues, it

can only be that the supply is regenerated as fast as it is utilized.

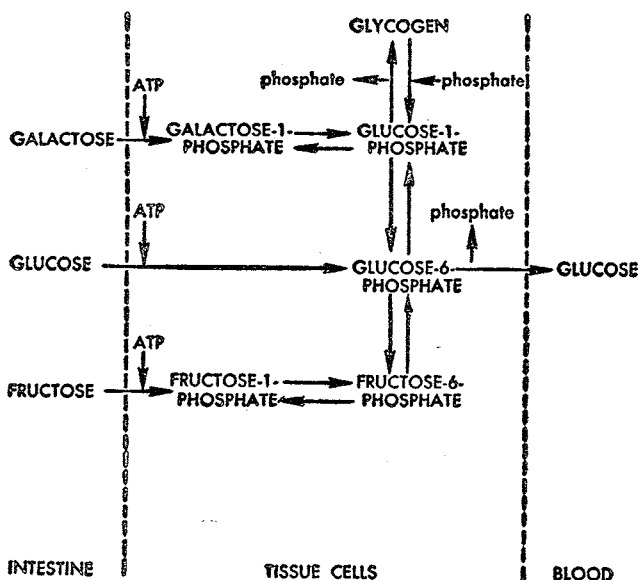


Figure 52. *ATP and Glycogen*

This means that the body must somehow succeed in building up molecules of ATP out of ADP as fast as ATP molecules are stripped to ADP in a myriad coupled reactions. We would therefore seem to be worse off than ever. Having solved the problem of moving glucose (and other compounds) uphill to the tune of 4 kilocalories per mole at the expense of hydrolyzing ATP to ADP, we must ask tissue to solve the problem of moving ADP back uphill to ATP, this time supplying 8 kilocalories per mole.

Yet this is exactly what is done in the process of anaerobic glycolysis, which I discussed in the previous chapter.

Throughout the 1930s, biochemists were isolating intermediates in glycolyzing yeast or tissue, and then finding enzymes that would catalyze reactions involving those intermediates. By identifying the intermediates and studying the

reactions, they were able to put together the entire catabolic chain from glucose to lactic acid, after the fashion of a jigsaw puzzle.

First glucose has a phosphate group attached to it and then, after conversion to fructose, a second phosphate is attached. (The fructose diphosphate thus formed is the Harden-Young ester I mentioned in the previous chapter.) Each phosphate attachment requires the services of ATP, so it would seem that the glycolytic chain is just aggravating the problem, using up ATP instead of forming it. But we are not through.

The six-carbon fructose diphosphate is split up into two three-carbon halves ("glyceraldehyde-3-phosphate"), each carrying one of the phosphates. What happens to this compound is most easily explained in Fig. 53.

As you see, a second phosphate group is added to it, but not by means of an ordinary condensation. That is, the elements of water ($\text{H}-\text{O}-\text{H}$) are not split out. Instead, two hydrogen atoms are split out and nothing more; a process called "dehydrogenation." (These hydrogen atoms do *not* come off as hydrogen gas, although in the equation I have written the hydrogen simply as 2H . Actually their fate is rather complicated, but I shall leave this for discussion later in the book.)

The result of this dehydrogenation is that an acyl phosphate is formed, which, as I explained earlier in the chapter, is high-energy. But where did the energy come from for the formation of such a high-energy phosphate bond? Why, from the loss of the two hydrogen atoms. Dehydrogenation involves a decrease in free energy of from 35 to 70 kilocalories per mole and this is far, far in excess of the 8 kilocalories required to form the high-energy bond. (Under strictly anaerobic conditions, however, this doesn't represent the waste it seems to, because the two hydrogen atoms are returned, with their excess energy, at a later stage.)

Later in the chain a second high-energy bond is formed by a reaction that is less energetic than that of a dehydrogenation, but is nevertheless sufficient for the conversion of an already present low-energy phosphate bond into a high-energy one.

Any compound containing a high-energy phosphate bond (for example, 1,3-diphosphoglyceric acid) can react with ADP to transfer that phosphate bond to the latter compound,

forming ATP, with no energy to speak of being required for the transfer. The formation of a high-energy phosphate bond of any kind, therefore, is equivalent to the formation of a molecule of ATP.

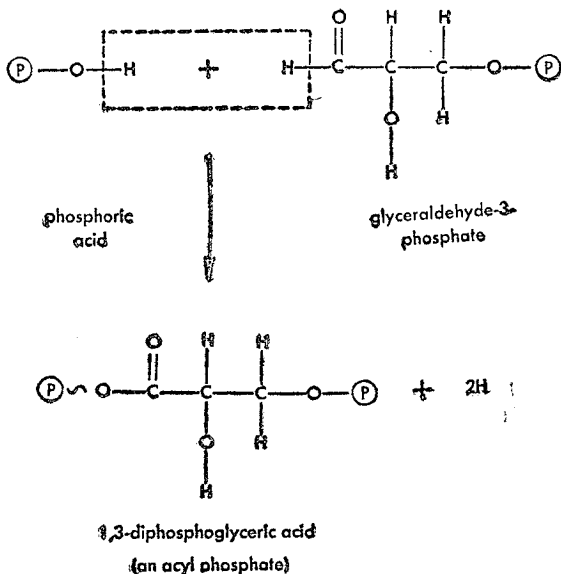


Figure 53. *Dehydrogenation*

We can summarize then as follows: In anaerobic glycolysis, each glucose molecule, in its conversion to lactic acid, begins by making use of two ATP molecules. The glucose, however, is eventually converted to two three-carbon compounds, each of which delivers two molecules of ATP to the body by the time it is converted to lactic acid.

The six carbons of glucose, therefore, use up two ATP molecules and form four, all together, for a net production of two (see Fig. 54), which may then be used in coupled reactions to make anabolism possible.

But glycolysis, as I said in the previous chapter, is an inefficient way of getting energy. To live on glycolysis alone would leave any organism as torpid as yeast cells. For some-

thing better, we must turn to those processes that involve oxygen.

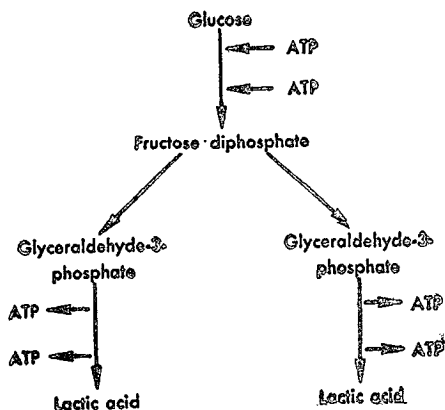


Figure 54. *Schematic Diagram of Glycolysis*

Chapter 22

THE TRANSFER OF ELECTRONS

The lion's share of the energy produced in metabolism is obtained from reactions that ultimately involve atmospheric oxygen. In order to speak freely about these reactions, however, there is some preliminary material that should be discussed first.

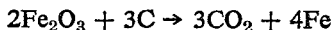
The common term for the burning of any inflammable object is "combustion," from a Latin word meaning "to burn up." Once Lavoisier discovered that the combustion of wood involved the combination of its components with oxygen, the more descriptive word "oxidation" suggested itself at once.

Oxidation turned out to be not only a more descriptive word than combustion, but also a more general word. The

slow combination of metal with oxygen is what forms a "calx" (or what we today would call an "oxide"), as Lavoisier also showed, and although that is not combustion as that word is generally understood, it is an oxidation. In other words, oxidation came to mean any combination with gaseous oxygen, whether a flame and outright burning was involved or not. And oxygen, bringing about an oxidation as it does, is an example of an "oxidizing agent." (It may seem tautological to speak of oxygen as an oxidizing agent, but it will be explained soon enough that oxygen is not the only substance that will bring about oxidation.)

On the other hand, it was long fashionable to speak of an ore as being "reduced" to a metal. "Reduce" is derived from Latin words meaning "to lead back," and if iron, for instance, is observed to turn to rust (and the most common of the iron ores is a form of rust, compacted to rocklike hardness) it is natural to think of rust being "led back" to iron.

It is common to be working with a metal oxide as an ore, and to form the metal it is necessary to remove the oxygen atom or atoms. Thus, if iron ore is heated with coke, the carbon atoms of the coke combine with the oxygen atoms of the iron ore, leaving metallic iron behind:



Since carbon brings about the reducing, carbon is an example of a "reducing agent."

Oxidation, then, is the addition of oxygen to a substance and reduction is the removal of oxygen from a substance. The two are clearly opposite in nature. Oxidation is the reverse of reduction; reduction the reverse of oxidation.

Another common way of removing oxygen from combination with a metal is to treat the oxide with hydrogen. Hydrogen, like carbon, is a reducing agent. It will combine with oxygen and will, for instance, restore copper, after it has been heated to a dark tarnish, to its original ruddy shine.

Organic chemists were particularly interested in reactions involving hydrogen. They found that certain chemicals, rich in oxygen, which had the capacity of tarnishing metals by formation of oxides, had also a tendency to remove hydrogen atoms from the molecules of organic compounds. These

oxygen-rich chemicals, as for instance potassium permanganate (KMnO_4), and potassium chromate (K_2CrO_4), were, therefore, oxidizing agents. They seemed to be adding some of their rich supply of oxygen to metals in tarnishing them. They also seemed to be using that oxygen supply for combination with hydrogen, removing the latter from the organic molecule and forming water. The loss of hydrogen ("dehydrogenation") could therefore be viewed as a kind of oxidation in itself, even if no oxygen atom were directly added to the organic molecule under consideration.

On the other hand, if an organic molecule, oxidized in this fashion, is treated with hydrogen gas (usually in the presence of some catalyst) the oxidation reaction can be reversed. The addition of hydrogen ("hydrogenation") could therefore be viewed as a reduction even if, again, no oxygen molecule is involved.

In short, the organic chemist came to consider oxidation as the addition of oxygen *or* the removal of hydrogen; while reduction was the removal of oxygen *or* the addition of hydrogen.

Even this is not general enough. Metals will tarnish under the influence of fluorine, chlorine, bromine, and sulfur in a manner irresistibly reminiscent of oxidation, even though neither oxygen nor hydrogen is involved. In the same way, metallic sodium, potassium, magnesium, and calcium will act as reducing agents, even where neither oxygen nor hydrogen is involved. It would seem clumsy, however, to extend the definitions of oxidation and reduction to cover addition or removal of an indefinite number of elements, and something more general was sought. The greater generalization only came with the discovery of the electron.

In the twentieth century, it came to be realized that whenever a substance underwent an oxidation it lost electrons. When a substance was reduced it gained electrons. It seemed natural, then, to remove oxidation and reduction from any connection with specific elements and to define oxidation as the loss of electrons and reduction as the gain of electrons.

For instance, when sodium metal reacts with chlorine gas to form sodium chloride, the sodium atom loses an electron to form sodium ion. Sodium is therefore oxidized by the chlorine; and chlorine is an oxidizing agent. On the other hand, in this same reaction chlorine atoms gained electrons to become chloride ions. Chlorine is therefore reduced by

the sodium; and sodium is a reducing agent (see Fig. 55).

Again, if carbon is burned in oxygen, carbon dioxide is formed, with each carbon atom being bonded covalently to two oxygen atoms. No electrons are actually transferred, none are completely lost to any atom or completely gained by any atom. However, the oxygen atom has a stronger hold on the shared electrons than has the carbon atom. The carbon atom loses those electrons, *in part*, and is oxidized, oxygen being the oxidizing agent. On the other hand, the oxygen atom gains those electrons in part, and is reduced, carbon being the reducing agent.

As you see, it is impossible to have an oxidation all by itself or a reduction all by itself, from this point of view. Electrons, in ordinary chemical reactions, cannot exist free, but can only be transferred from one atom to another, wholly or in part, the loser being oxidized, the gainer reduced. In this view, the substance being oxidized is always a reducing agent for the other substance involved, which, being reduced, is in its turn the oxidizing agent for the first substance. (Of course, if electrons are shared equally, as when atomic chlorine reacts to form molecular chlorine, the result is neither an oxidation nor a reduction.)

Because of this inextricable binding of oxidation and reduction, reactions involving them are called "oxidation-reduction reactions," a term that is sometimes shortened, rather uneophoniously, to "redox reactions."

How do the processes of hydrogenation and dehydrogenation fare under the new electronic dispensation?

The effect of the loss of hydrogen must be weighed according to whether the loss of electrons is also involved, since it is the fate of the electrons that counts in this definition. Sometimes a compound is capable of releasing a hydrogen atom *without* the electron that "rightly" belongs to it. The hydrogen, minus its electron, is a positively charged hydrogen ion (H^+). What is left of the compound is a negatively charged ion, and the process is called "ionization." Has the compound gained an electron in this way and has it been reduced?

The answer is, no. A hydrogen ion is released only when its hold on the electrons it shares (generally with oxygen) is, for one reason or another, already so weak that it "falls off" the compound. It has not lost electrons in the process of ionization, nor has the rest of the compound gained them.

Rather, the ionization is a symptom of the fact that electrons have already been transferred at some previous stage

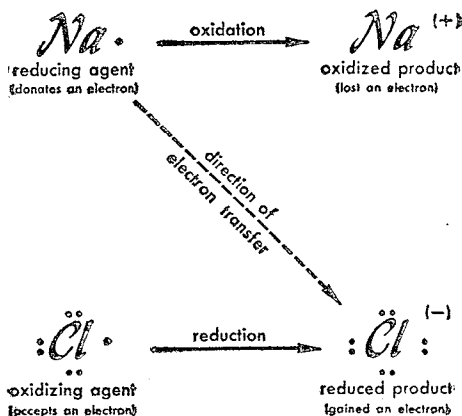
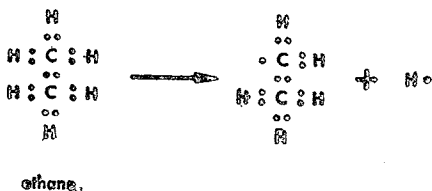


Figure 55. Oxidation-Reduction Reaction

in the formation of the compound. Ionization itself, therefore, is neither oxidation nor reduction!

But suppose a hydrogen atom is removed from a compound, along with its electron. The situation is presented in Fig. 56. (The question of what happens to the hydrogen atom will be considered later. For now, we shall just let it hang around loose, so to speak.)

Here the organic compound may be considered to have full ownership of one electron where it had previously had part ownership of two. Since the carbon atom has a slightly stronger hold on electrons than the hydrogen atom has, car-



ethane.

Figure 56. Electronic Dehydrogenation

bon's part ownership of two amounts to a trifle over one full electron. Changing that "trifle over one" for merely "one" involves a loss of a part of an electron and so the compound can be said to have been oxidized. As you see, dehydrogenation remains oxidation even under the new electron definition.

By a similar argument, we can show that the loss of an oxygen atom (either by itself or as part of a hydroxyl group) *with* its electrons gives the remainder of the compound full control of one electron where previously it had a less-than-half share of two. The compound has gained part of an electron so that removal of oxygen represents a reduction.

Consequently, if any compound loses both a hydrogen atom and a hydroxyl group ("dehydration"), or gains both a hydrogen atom and a hydroxyl group ("hydration"), the oxidation effect of one cancels the reduction effect of the other. Consequently neither dehydration nor hydration is either an oxidation or a reduction. This is also true if two compounds are involved rather than one, so that neither hydrolysis nor condensation is either an oxidation or a reduction.

But let's go back to the dehydrogenation as pictured in Fig. 56. Notice that after the removal of the hydrogen atom the hydrocarbon is left with an unpaired electron. Any group of atoms containing an unpaired electron is called a "free radical."

A free radical generally has only a fugitive existence, for the unpaired electron represents a highly unstable situation. It seeks to make a pair of itself, so to speak, and a chemical reaction of some sort quickly takes place.

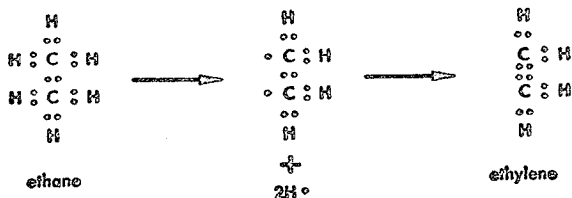
Of course, if two hydrogen atoms (each with its electron) are removed, so that two unpaired electrons are formed, these two can then pair with each other to form a "double bond," as in Fig. 57.

Most oxidations of organic compounds do indeed involve such a double dehydrogenation; the loss of two hydrogen atoms. Such a dehydrogenation involves a large loss of free energy; one so large, in fact, that such reactions, taking place during catabolism, are *the* way in which living tissue obtains energy for the formation of high-energy phosphate bonds.

Nevertheless, such dehydrogenations do *not* take place under ordinary circumstances. An organic compound can exist

for years without any measurable tendency to dehydrogenate, despite the large negative free energy change involved.

Electronic



Kekule

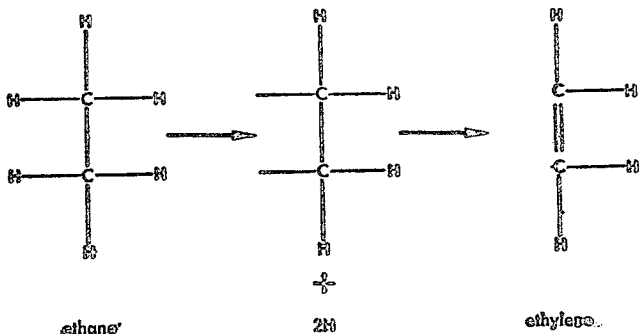


Figure 57. Double Bond Formation

Clearly, a large energy of activation must also be involved. In 1935, Michaelis (of the Michaelis-Menten equation) pointed out that it was very unlikely that any organic compound would lose two hydrogen atoms simultaneously. It was much more likely that it would lose first one and then another. There would therefore be an intermediate stage at which only one electron was gone and the compound would then be a free radical.

Now the free radical is unstable and forming it requires a large input of free energy. The loss of the second hydrogen atom afterward more than makes up for that; far more.

Nevertheless, that first electron must go first and the energy for that reaction must be found. That energy is the energy of activation, and it is the difficulty of forming the intermediate free radical that allows organic compounds (including our own precious bodies) to remain immersed in oxygen indefinitely without rapid over-all conversion to carbon dioxide and water.

Yet there are enzymes in tissues that catalyze dehydrogenations at body temperature. This is done by lowering the energy content of the free radical; stabilizing it, in other words. The energy of activation is thus lowered to the point where the kinetic energy of molecules at body temperature is sufficient to bring about free radical formation. Once that happens, of course, the loss of the second hydrogen and the formation of the double bond proceeds at once.

But how is the free radical stabilized?

Well, throughout the nineteenth century chemists tried to produce substances that we would now call free radicals under conditions of more than fugitive existence, but they always failed. In 1900, however, the American biochemist Moses Gomberg finally succeeded. He showed that certain complicated free radicals could be made to exist for in-

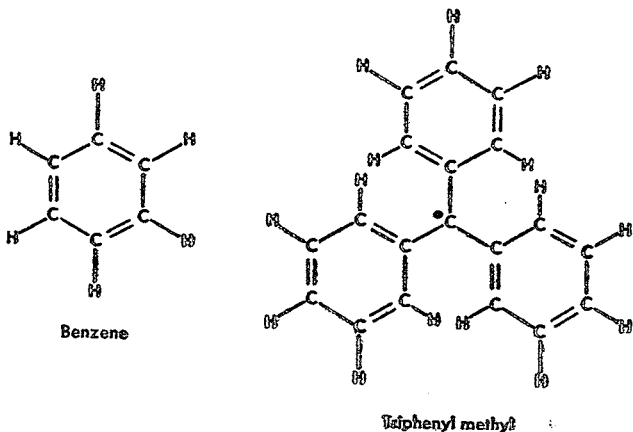


Figure 58. *A Free Radical, Triphenyl Methyl, and Its Relation to the Benzene Ring*

definite periods and to show their existence by visible changes in the colors of solution.

The free radical first produced in this manner ("triphenyl methyl") consisted of a carbon atom attached to three "benzene rings." (Benzene itself consists of six carbon atoms and six hydrogen atoms in a symmetrical circle.) All three benzene rings, we now know, exist in a single plane and are symmetrically disposed about the central carbon atom, as in Fig. 58. It is the central carbon atom, shown with only three Kekule bonds, and carrying an unpaired electron (symbolized by the dot) in place of the fourth, that makes the whole molecule a free radical (see Fig. 58).

But why should this particular free radical be so stable? This was first explained reasonably well by the American chemist Linus Pauling in the early 1930s. Pauling took advantage of theories developed in the 1920s to the effect that the electron was not merely a featureless little ball, but a collection of wave forms that could be smeared fuzzily over a larger or smaller volume of space, depending on circumstances.

Pauling developed what is called "resonance theory," using mathematics based on the wave nature of electrons, to show, among other things, that the more an unpaired electron could be fuzzily spread out, the more stable it could be made to be.

The spreading out of an electron could be made more likely if the molecule of which it forms a part exists in a single plane and is highly symmetrical. Gomberg's triphenyl methyl lives up to these requirements admirably. Although I have drawn the unpaired electron as a single dot in the diagram, this does not represent the facts as chemists now believe them to be. Rather it should be conceived as a kind of fog, enclosing the molecule, with condensations of it symmetrically placed here and there.

Michaelis, with the theory of resonance fresh in the air, pointed out that an enzyme combined with a substrate might well form a far more symmetrical system than the substrate alone. The enzyme-substrate free radical would therefore be more stable than the substrate free radical itself, as the unpaired electron spreads out far more eagerly in the former case. Less energy is required to form the enzyme-substrate free radical than would be required to form the substrate

free radical itself, and the net effect is that the enzyme successfully catalyzes the reaction.

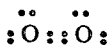
Now we ask what happens to the two hydrogen atoms removed from an organic compound by dehydrogenation. One possibility is that the two may combine to form molecular hydrogen, which will be given off as a gas. This, however, is not observed to happen in the dehydrogenation of an organic compound by any organism other than a very few kinds of bacteria.

The alternative is that of having the atoms transferred to an "acceptor" molecule which, through the addition of the hydrogens, with their electrons, is itself reduced. The most obvious acceptor molecule is oxygen and we should, therefore, take a look at the electronic structure of that element.

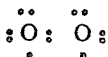
The oxygen atom contains eight electrons, two in an inner shell and six in an outer, so that the atom may be symbolized electronically as:



The atom does not exist free under ordinary conditions, but exists in pairs, forming oxygen molecules. The natural assumption would be that, in doing so, it forms a pool of two pairs of electrons so that each atom has the necessary eight electrons in its outer shell, thus:



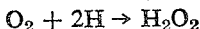
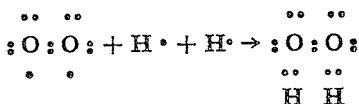
The trouble with this assumption is that oxygen seems to possess strong magnetic properties, and chemists are certain that such properties go along with the existence of unpaired electrons. The electronic formula of the oxygen molecule is pictured thus, then:



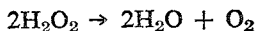
This is a sort of double free radical that is, nevertheless,

stable; a quite exceptional case that is of interest to theoretical chemists but need not detain us.

Such an oxygen molecule can easily accept two hydrogen atoms in the following fashion (the reaction being represented both electronically and in ordinary fashion):



The compound H_2O_2 is "hydrogen peroxide," a rather unstable compound with a tendency to break down to water and oxygen, thus:



The breakdown of hydrogen peroxide is a spontaneous reaction that represents a decrease of free energy (see Fig. 59). That it happens as slowly as it does (and hydrogen peroxide, if kept in dilute solution at a fairly low temperature, can be stored without much loss for considerable periods of time) is owing to the fact that the O—O bond must first be broken, and that this naturally represents a considerable energy of activation.

Even gentle heating is enough to supply this energy of activation and to accelerate the hydrogen peroxide breakdown considerably. A number of chemical substances will catalyze the breakdown; iron filings will, for instance. So will an enzyme of widespread occurrence called "catalase."

Since the passage from hydrogen peroxide to water involves a decrease in free energy, one can say that the passage from oxygen to water involves a greater decrease in free energy than the passage from oxygen to hydrogen peroxide. Should we not expect, then, that oxygen in accepting hydrogen would slide down the steeper slope and form water every time?

The fly in the ointment is the energy of activation involved in breaking the O—O bond. When oxygen accepts hydrogen at high temperatures (as when a hydrogen-oxygen flame is

ignited, or when an organic compound burns) the oxygen-oxygen bond is broken without trouble and water is formed.

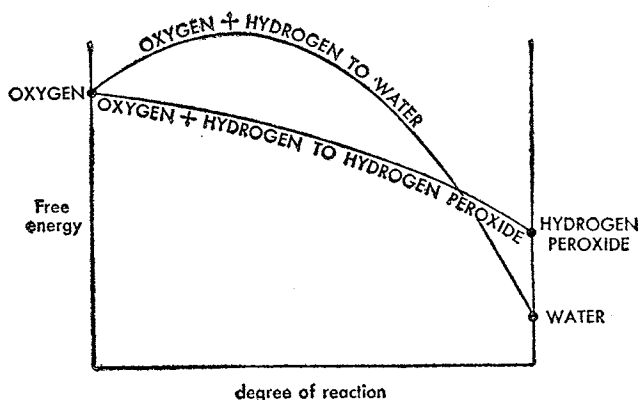


Figure 59. *Hydrogen Peroxide and Water*

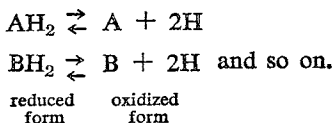
When the oxygen accepts hydrogen at room temperature, as it does in some enzyme-catalyzed reactions, energy may be lacking to break the oxygen-oxygen bond and hydrogen peroxide is formed.

Nevertheless, hydrogen peroxide is not to be found in living tissue in perceptible concentration—and a good thing, too, for it is highly toxic. One possible explanation for this rests with the catalase that *is* present. Any hydrogen peroxide formed would be promptly broken down to water through its action. Nevertheless, as we shall see, there is reason to think that hydrogen peroxide is not formed to any large degree in the first place, and catalase may be only a useful watchdog on whose services it is not necessary to call frequently.

While oxygen is a likely and logical acceptor of hydrogen atoms, it is not the only possible one. Hydrogen atoms being donated by one organic compound may be accepted by a variety of other organic compounds.

To see how this can be, consider that many organic compounds can exist in a reduced form and an oxidized form, with the difference lying in the presence and absence of

two hydrogen atoms respectively. We can represent this as follows:



Different compounds will naturally have different oxidation-reduction equilibria. For instance, the first equation listed above may have an equilibrium point far to the right, so that AH_2 would have quite a tendency to lose two hydrogen atoms. On the other hand, the second equation could have an equilibrium point far to the left, so that B would have a tendency to gain hydrogen atoms.

If, then, AH_2 and B were mixed the following reaction would take place:



AH_2 , losing hydrogens, would be oxidized in the process by B, the oxidizing agent. B itself, gaining hydrogens, would be reduced in the process by AH_2 , the reducing agent.

Now all we need do is devise some method that will measure this tendency of hydrogen to flow from one compound to another so that we can predict, perhaps, which compound will oxidize (or reduce) which.

To do this, it is necessary to remember that the donation and acceptance of hydrogen atoms is really a specialized aspect of the donation and acceptance of electrons. Well, any compound that has a tendency to give up electrons sets up a "pressure" tending to drive them through an electric circuit. Such pressure is the electric potential. By setting up chemical cells of the type I talked about in Chapter 10, the electric potentials associated with a particular oxidation reaction can be measured as so many volts. Adjusted to certain standardized conditions, they become "oxidation potentials."

The oxidation potential of the reaction in which a hydrogen atom gives up an electron to become hydrogen ion is arbitrarily set equal to zero. Any atom or molecule with a smaller tendency to give up an electron has a positive oxida-

tion potential. An atom or molecule with a greater tendency to give up an electron than is true of hydrogen has a negative oxidation potential.

Any substance will accept electrons (and, on occasion, the associated hydrogen atom) from another substance with a less positive, or negative, oxidation potential. Any substance will donate electrons (and, on occasion, the associated hydrogen atom) to another substance with a more positive, or less negative, oxidation potential.

It is even possible to visualize a series of oxidation-reduction reactions in which the same electrons, or the same hydrogen atoms, are passed from compound to compound down the line of decreasingly negative and increasingly positive oxidation potentials.

That, as a matter of fact, is precisely what takes place in living tissue and it is to living tissue we must now return.

Chapter 23

LIFE WITH AIR

In considering reactions involving atmospheric oxygen, it would seem natural to study the manner in which oxygen is absorbed by living tissue itself (as contrasted with its merely being taken into the lungs).

Slices of tissue can be obtained, from sources as far apart in the realm of life as the potato and the mouse, in which the enzyme systems are sufficiently intact to continue catalyzing those reactions that utilize oxygen exactly (it is presumed) as they do in actual life. For some purposes, intact cells are not needed, as oxygen can be taken up by reactions catalyzed by relatively pure enzyme solutions.

Such reactions are usually conducted in a small closed container so that pressure changes in the limited air supply are large enough to be followed. The container consists of two parts, one in which the substrate solution is placed and another in which the enzyme solution or the tissue slices

are to be found. A gentle tipping back and forth would mix the two and start the reaction at a given moment.

In the ordinary course of affairs, oxygen would then be taken up and carbon dioxide would be given off. The uptake of oxygen would tend to produce a partial vacuum in the container, but the production of carbon dioxide (which is sure to take place if tissue slices are used) would tend to make up for that and maintain normal gas pressure. If, however, the vessel also contained a small recess in which might be placed some solution that absorbed carbon dioxide rapidly, the partial vacuum produced by oxygen uptake would *not* be balanced. The fall in pressure would then indicate consumption of oxygen.

The container is connected by a tight seal to a U-tube, open to the atmosphere, with a colored liquid at the bottom. On one side of the U-tube the colored liquid would be under the pressure of the air within the container; on the other side it would be under ordinary atmospheric pressure. As the enzyme-catalyzed reaction in the container proceeds, the gas pressure within the container decreases with oxygen uptake, and the atmospheric pressure overbalances it. The liquid moves downward on the side of the U-tube open to the air and up on the side connected to the container.

Naturally, precautions are taken to make sure that the movement of the liquid is only in response to the partial vacuum within the container. The container is kept in a water bath held at constant temperature to avoid changes in air volume with temperature change. A container minus enzyme is simultaneously run so as to record changes that are due entirely to the continuing small alterations in atmospheric pressure.

In the end, the shifting level of the liquid in the U-tube can be made to record, with great delicacy, the rate at which oxygen is taken up as the result of enzyme-catalyzed reactions. A device measuring gas pressure is called a "manometer" from Greek words that may be rather freely translated as "gas-measure." Since the manometer just described was designed in its most popular form in 1923 by the German biochemist Otto Warburg, it is widely known as a "Warburg manometer."

Now if minced muscle is tested in a Warburg manometer, oxygen is taken up as the muscle enzymes catalyze oxygen-

consuming reactions. With time, however, the rate of oxygen uptake slowly drops off and declines to low values. It might seem as though the enzymes involved have been denatured and have lost their catalytic ability. However, this loss of activity takes place even when the reactions are conducted under conditions where one would expect the enzymes not to be denatured. A second possibility is that the substrates upon which the enzymes were exerting their catalytic influence have been used up.

In the latter case, oxygen uptake might be started once more if additional substrate were added to the mixture. It remained only to decide which compounds might be considered substrates. One guide to the answer lay in the fact that certain enzymes, which had one specific compound or another for their substrates, had been isolated from muscle tissue. It seemed logical to suppose this compound was the substrate within tissue as well as within a test tube, so such compounds were added to the minced tissue.

In 1935 the Hungarian biochemist Albert Szent-Györgyi found that oxygen uptake could be restored to its original levels if, to the minced tissue, any one of four compounds—succinic acid, fumaric acid, malic acid, or oxaloacetic acid—was added. These are all similar in molecular structure as you can see in the formulas given in Fig. 60. All, for instance, are “dicarboxylic acids,” since each contains two carboxyl groups (COOH) per molecule.

Apparently each compound is somehow involved in the reaction or reactions that consume oxygen. However, none of the compounds can be involved all by itself because the amount of oxygen that is consumed after the addition of one of the compounds is far greater than that required to consume all the quantity added. Clearly, what happens is that a series of reactions is involved in which the compound (succinic acid, for instance) is regenerated as fast as it is consumed, so that it exerts what might be considered a catalytic influence. As the reaction series continues, succinic acid is not regenerated with perfect efficiency so that there is slow wastage and, eventually, more succinic acid must be added if the series of reactions is to continue.

It can further be seen that the four compounds are not involved in four different series of reactions but in the same one. The proof of that is presented us by malonic acid.

If succinic acid is added to the tissue to start oxygen

consumption, the subsequent addition of malonic acid promptly stops it again. This is to be expected, since malonic acid is known to be a competitive inhibitor of succinic acid

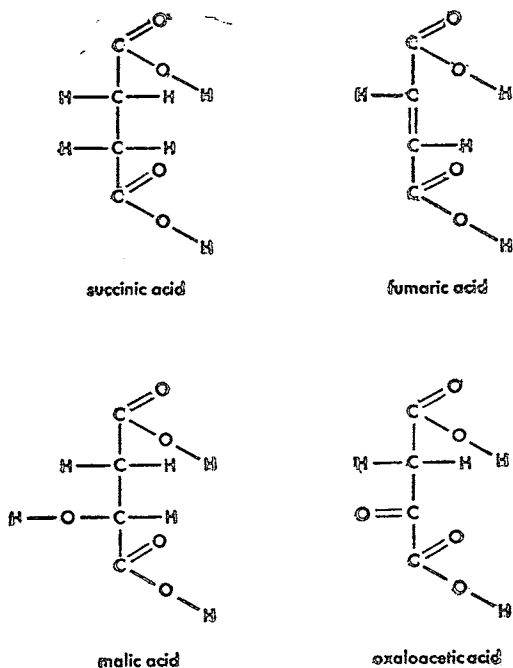


Figure 60. *The Dicarboxylic Acids*

(see Chapter 18). However, malonic acid also stopped oxygen uptake when it was added after the addition of any of the other three compounds. Now malonic acid does not directly inhibit the reactions involving malic acid or oxaloacetic acid, for instance. That it does so when added to minced tissue can only be because those reactions are part of a series that include succinic acid and, by putting a spoke in the succinic acid portion, malonic acid stops the whole series before and after.

After this much had been worked out, the German biochemist Hans Adolf Krebs went on to show that other com-

pounds, notably citric acid (which contains six carbon atoms and three carboxyl groups, rather than four carbon atoms and two carboxyl groups, as did Szent-Györgyi's compounds), could also start up the oxygen-consuming series of reactions. By 1940 he was able to build up a coherent scheme in which all the necessary compounds were placed in logical sequence, and that scheme has had to be modified in only minor fashion since.

This series of reactions is commonly called the "Krebs cycle" for obvious reasons, or, more descriptively, the "citric acid cycle" or "tricarboxylic acid cycle," since citric acid possesses three carboxyl groups.

Making use of as few formulas as we can and going into as little detail as possible, let's see how the Krebs cycle works.

Anaerobic glycolysis, as I explained in Chapter 20, converts a molecule of glucose to two molecules of lactic acid, with the net production of two high-energy phosphate bonds in the ultimate form of ATP. It is the lactic acid now that must be catabolized further with the formation of additional high-energy phosphate bonds.

The process begins with the removal of two hydrogen atoms from lactic acid (dehydrogenation), forming "pyruvic acid" as shown in Fig. 61.

Now a $C=O$ combination is, by itself, called a "keto group" and a $COOH$ is the (by now familiar, I hope) carboxylic acid group. Any compound which, like pyruvic acid, contains both is called a "keto acid."

In the body, keto acids almost invariably undergo a reaction that reduces them in size by one carbon atom. We can leave out some of the finer details and say that the net effect in this particular case is to convert pyruvic acid (three carbon atoms) to acetic acid (two carbon atoms) as shown in Fig. 61.

This is an example of an "oxidative decarboxylation"; oxidative, because two hydrogen atoms are removed, and decarboxylation, because carbon dioxide is also removed. This particular oxidative decarboxylation is catalyzed by an enzyme that makes use of a coenzyme that in turn contains an atom grouping that makes up a rather complex molecule called "thiamine." This, as it happens, is vitamin B_1 , the absence of which produces beriberi in man. It was the ef-

fect of the lack of this vitamin that Eijkman studied in the 1890s (see Chapter 18) and it was those studies that initiated

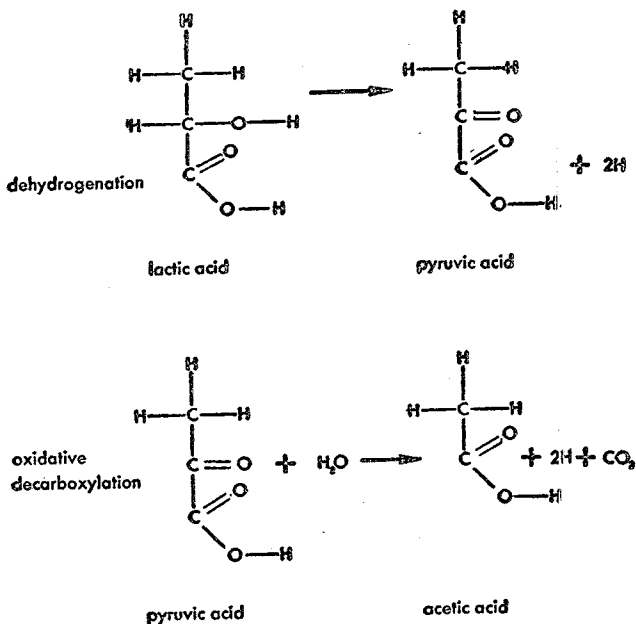


Figure 61. *Dehydrogenation and Oxidative Decarboxylation*

modern knowledge of vitamins generally.—And it is at this point that the vitamin exerts its effect, for without it the pyruvic-to-acetic reaction halts and all the vast interlocking of metabolic reactions slows to grinding inefficiency, like a gear with a broken tooth.

In the conversion of organic substances to water and carbon dioxide, the water is formed by dehydrogenation reactions, for the hydrogen atoms given up by a compound like lactic acid are eventually combined with oxygen. The carbon dioxide is formed by oxidative decarboxylations of keto acids, for the most part.

It is the dehydrogenation step that produces the energy utilized by the body. The mere elimination of carbon diox-

ide from a compound does not produce the energy necessary for the formation of high-energy phosphate bonds. Consequently, we can say that the body obtains its energy by burning hydrogen, and that the burning of carbon is only incidental. This is not surprising, as I have already explained (see Chapter 8) that the burning of hydrogen liberates far more heat per unit weight than does the burning of carbon.

In the next step, acetic acid is added to oxaloacetic acid and in this way enters the Krebs cycle (of which, you will remember, oxaloacetic acid was one of the first components discovered).

During the early 1940s the details of this addition were unknown. In 1947, however, Lipmann (who had introduced the high-energy phosphate bond) isolated a compound that served as coenzyme in this reaction which he called Coenzyme A (A for "acetic acid").

The molecule of Coenzyme A is quite a complicated one and I won't bother you with it. A large portion of it consists of an atom grouping called "pantothenic acid," which the body cannot manufacture for itself and which must therefore be present as such in the diet. Pantothenic acid, like thiamine, is one of the B vitamins, in consequence.

The reaction portion of the Coenzyme A molecule is a

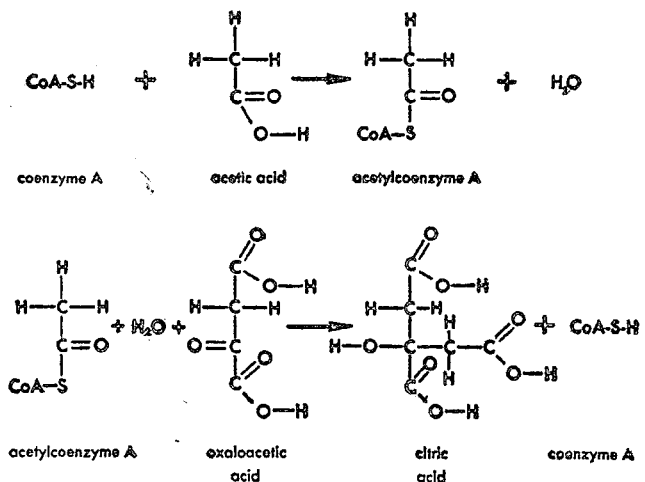


Figure 62. Coenzyme A

thiol group (—S—H). Consequently, the Coenzyme A molecule is often represented by the shorthand symbol: CoA—S—H .

Coenzyme A will condense with acetic acid as in Fig. 62 to form "acetylcoenzyme A." In fact, it takes part in the reaction in which pyruvic acid is oxidatively decarboxylated so that acetic acid itself never has a chance to form. It is acetylcoenzyme A from the start.

The portion of the acetic acid that is attached to the Coenzyme A molecule is sometimes called the "two-carbon fragment," a name dating back to the days when its exact nature—other than that it contained two carbon atoms—was uncertain.

There are many reactions in the body that involve the transfer of a two-carbon fragment from one compound to another, and in every case Coenzyme A is the workhorse that does the transferring. (Later, Coenzyme A was found to be part of a reaction involving a compound called "alpha-ketoglutaric acid," another member of the Krebs cycle, and here its job was to transfer a four-carbon fragment. Coenzyme A is not to be confined to two-carbon fragments only and, in fact, we shall come back to it later in the book, in order to describe still other tasks of transferral with which it is concerned.)

The addition of acetylcoenzyme A to oxaloacetic acid is also shown in Fig. 62. This complicated reaction (but if you follow it through carefully, you will see that every atom on the left is accounted for on the right) results in the formation of the six-carbon citric acid, while Coenzyme A is reformed and can rush back, so to speak, to pick up another molecule of acetic acid for transfer purposes.

Once citric acid is formed, it proceeds to undergo a series of reactions which, in the end, strip off the two-carbon fragment that was added to form it; thus reconverting the citric acid to oxaloacetic acid again. The oxaloacetic acid can now pick up a new two-carbon fragment, which will be stripped off in another turn of the Krebs cycle; whereupon it will pick up still another and so on indefinitely.

The formulas I have already given you in this chapter illustrate the key points in the process: the oxidative decarboxylations, the dehydrogenations, the catalyzed transfers by coenzymes, and so on. We can pass on, then, to present, without further formulas, a diagrammatic scheme of the

manner in which lactic acid is catabolized. This is presented in Fig. 63, which, I must warn you, hits only the high points.

As you see, lactic acid begins by undergoing one dehydrogenation in its conversion to pyruvic acid, a second dehydrogenation in the formation of acetylcoenzyme A, and four more dehydrogenations in the course of the Krebs cycle. That's six dehydrogenations all together, representing a loss of twelve hydrogen atoms.

Of course, the lactic acid molecule contains only six hydrogen atoms, as you will see if you check back to the formula on p. 316. How, then, can it lose twelve?

Well, as the summary of the Krebs cycle indicates, three different water molecules enter it at various points. They supply the remaining six hydrogen atoms.

In addition, the series of changes undergone by lactic acid and the compounds to which it is converted involve a de-

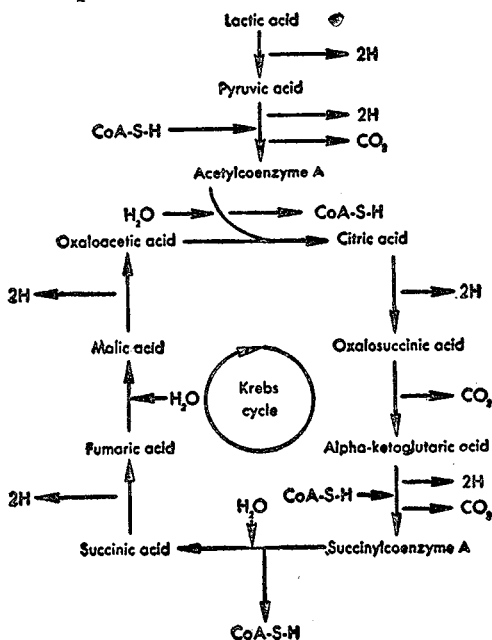
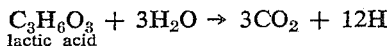


Figure 63. *The Catabolism of Lactic Acid*

carboxylation during the formation of acetylcoenzyme A and two more decarboxylations in the course of the Krebs cycle, for a total formation of three molecules of carbon dioxide.

The over-all equation, then, is:



But there are some questions that might arise now. First, let's get back to the matter of the fate of the hydrogen atoms. Clearly, oxygen must be involved since, after all, the workings of the Krebs cycle were determined originally by oxygen uptake experiments. Yet oxygen doesn't accept the hydrogen atoms directly, for no hydrogen peroxide is formed in tissue experiments even when tissues low in catalase are used.

Presumably, then, oxygen accepts the hydrogens indirectly, and in such a fashion as to produce water and not hydrogen peroxide. The obvious guess is that the hydrogen atoms are picked up by some acceptor other than oxygen, and then, ultimately, passed on to oxygen. Just as Coenzyme A transfers two-carbon fragments, so there must be coenzymes that transfer hydrogen atoms; "hydrogen carriers," in other words.

Actually, hydrogen carriers were discovered a generation before Coenzyme A was. The first coenzyme found to serve as a hydrogen carrier was, in fact, the very first one discovered: the cozymase isolated by Harden and Young (see Chapter 18). By the early 1930s the molecular structure of cozymase had been worked out. I shan't go into detail on that, any more than I did in the case of Coenzyme A, but again I shall mention the salient points.

First, it contains a five-carbon sugar molecule, two phosphate groups, and a nitrogen-containing ring of atoms, all arranged in a fashion that puts the molecule into a class of compounds called "nucleotides." The nitrogen-containing ring has the pyridine arrangement, as I mentioned in Chapter 18, so that the whole molecule is now called "diphosphopyridine nucleotide," and this is almost invariably abbreviated as DPN. In 1934, Warburg (of manometer fame) isolated a very similar coenzyme that differed only in that its molecule contained three phosphate groups rather than

two. It was inevitably named "triphosphopyridine nucleotide" and abbreviated TPN.

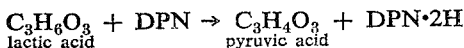
In both DPN and TPN the particular version of the pyridine ring is that of a compound known as "nicotinamide," which the body cannot form for itself from anything simpler than the very similar compound, "nicotinic acid." Nicotinamide or nicotinic acid must therefore be present in the diet as such, and they represent another of the B vitamin group.

Through the 1930s and 1940s increasing numbers of enzymes that catalyzed the dehydrogenation of one compound or another were isolated, and they were therefore termed "dehydrogenases." A large proportion of them were found to include either DPN or TPN as the necessary coenzyme. Because of the characteristic presence of the pyridine ring in the coenzymes, the dehydrogenases were lumped under the general term "pyridino-enzymes."

In the case of every pyridino-enzyme, it is the DPN or TPN that is the active group, the "cutting edge" so to speak. However, it is the amino acid portion of the enzyme (the "apoenzyme") that determines the specificity; that is, determines which particular compound can approach the coenzyme and serve as substrate. Thus, a dozen different pyridino-enzymes with a dozen different amino acid structures might catalyze the dehydrogenation of a dozen different compounds, with no one enzyme able to affect any of the eleven compounds other than its own substrate, and with all the enzymes using DPN, say, as substrate.

In fact, of the six dehydrogenations taking place in the course of lactic acid catabolism, five involve a pyridino-enzyme, each one different. (The sixth dehydrogenation is that of succinic acid and I'll come to that later.)

The function of the DPN or TPN is to accept the hydrogen atoms given up by the substrate. While the substrate is losing hydrogen and being oxidized, the DPN or TPN is gaining hydrogen and being reduced. The reduced form of DPN or TPN can most simply be represented (without going into detail of structure) as $DPN \cdot 2H$ or $TPN \cdot 2H$. Thus, the dehydrogenation of lactic acid to pyruvic acid, catalyzed by the pyridino-enzyme, "lactic acid dehydrogenase," can be represented as follows (without bothering to write the actual structural formulas of lactic acid and pyruvic acid, since you can find them on p. 316):



The $\text{DPN}\cdot 2\text{H}$ can then pass on the two hydrogen atoms to something else and be re-oxidized to the original DPN. It is now ready to take up two more hydrogen atoms, which it can pass on in turn, then take up two more and so on.

But what is the "something else" to which the $\text{DPN}\cdot 2\text{H}$ passes on the hydrogen atoms?

Oxygen?

Let us see. If a pyridino-enzyme is set to work dehydrogenating a substrate such as lactic acid, in the absence of other enzymes and substrates and in a nitrogen atmosphere, the reaction will not proceed. The DPN present will accept hydrogen atoms, to be sure, but there will be nothing for them to pass them on to. The total number of DPN molecules present is so small compared to the number of lactic acid molecules that all the DPN may be reduced without any perceptible oxidation of the lactic acid. Unless the $\text{DPN}\cdot 2\text{H}$ that formed can unload the hydrogen atoms and proceed to additional cycles of reduction and oxidation, the reaction won't move.

The addition of oxygen to the solution does not help, so it is clear that oxygen cannot act as hydrogen acceptor where a pyridino-enzyme is involved.

If, however, certain chemicals, such as one called "methylene blue," are added, even in the absence of oxygen the reaction will proceed. Methylene blue will be accepting the hydrogen atoms from $\text{DPN}\cdot 2\text{H}$. Methylene blue itself is blue in color as one would expect from its name, but reduced methylene blue is colorless. The progress of the reaction is thus made visible by the gradual decolorization of the methylene blue.

But though this is interesting, it is no real solution to our actual problem, for there is no methylene blue in living tissue. All we show in such experiments is that DPN and TPN can indeed pass on their hydrogens under certain circumstances, but upon what compound do they unload their hydrogen *in living tissue*?

Let us for a moment pass on, then come back to this question.

During the 1930s (and on occasions before that) yellow substances, called "flavins" from a Latin word meaning "yellow," were isolated from food and tissue. Their structure was worked out by the end of the decade, and they were found to contain a three-ring system which, as it turned out, the body could not manufacture itself. It had to be present in the diet in the form of "riboflavin," still another of the B vitamins.

Warburg showed that the flavins were associated with enzyme systems to which they imparted a yellow color, and the first such enzyme to be discovered is still called "Warburg's old yellow enzyme." Eventually two important coenzymes were discovered among the flavins; one is "flavin mononucleotide" (FMN), and the other "flavin adenine dinucleotide" (FAD).

Enzymes that make use of FMN or FAD are called "flavo-enzymes" and these turned out, like those that made use of DPN or TPN, to catalyze dehydrogenations. For instance, the dehydrogenation of succinic acid (the one dehydrogenation in the course of lactic acid catabolism which, as I mentioned earlier, does not involve a pyridino-enzyme) takes place under the catalytic influence of "succinic acid dehydrogenase" for which FAD is the coenzyme.

Again, it is the function of these coenzymes to accept hydrogen atoms from the substrate, to become FMN \cdot 2H or FAD \cdot 2H, and then to pass those hydrogen atoms on to something else.

And here, at last, something new is added. The flavin coenzymes can pass hydrogen atoms on to oxygen, forming hydrogen peroxide. (Dehydrogenases of the flavo-enzyme variety are therefore called "aerobic dehydrogenases" because they can carry on their catalytic activity in the presence of oxygen as the only hydrogen acceptor. Dehydrogenases of the pyridino-enzyme variety are called "anaerobic dehydrogenases" because they cannot.)

Thus, if a flavo-enzyme is added to a system containing a pyridino-enzyme and a substrate (which by themselves cannot react), the flavo-enzyme will accept the hydrogen atoms from the pyridino-enzyme and pass them on to oxygen. The flavo-enzyme itself might have no effect on the substrate directly, and the pyridino-enzyme itself might have no effect on the oxygen. The two together, however, catalyze the

dehydrogenation of the substrate and the reduction of oxygen to hydrogen peroxide as shown, schematically, in Fig. 64.

The transfer, as shown, is in three steps, each step representing a new oxidation-reduction reaction. As I explained at the end of the previous chapter, the hydrogen atoms move step-wise in the direction of a continually increasing oxidation potential.

But then why cannot the hydrogen atoms pass directly from $\text{DPN} \cdot 2\text{H}$ to oxygen? That would still be in the direction of increasing oxidation potential. Why is an intermediate system with an intermediate oxidation potential required? To make use of a metaphor, if gravity will help pull you downward as you walk down two steps, one at a time, surely it will do so if you jump both at once. (It may puzzle you why *increasing* oxidation potential should be equated with walking *down* stairs. The assignment of positive and negative values to oxidation potential is purely arbitrary and I, personally, wish the decision had been the other way around. In any case, the flow of hydrogen atoms may be in the direction of increasing oxidation potential, but it is also in the direction of decreasing free energy, and that's what counts.)

As usual, the necessity for an intermediate step must involve the way in which that step lowers an energy of activation. Apparently the body cannot supply the energy required to pass the hump involved in the direct passage of hydrogen atoms from an organic compound to oxygen. Instead it always makes use of an inorganic factor as intermediate.

It was discovered in the 1950s that flavo-enzymes contain metal atoms, whereas pyridino-enzymes do not. Succinic acid dehydrogenase contains iron atoms, for instance, as part of the enzyme. Other flavo-enzymes contain copper or molybdenum. These metals have variable valences and can thus transfer electrons. Iron can form ferrous ion (Fe^{++}) and ferric ion (Fe^{+++}). By accepting an electron, ferric ion can become ferrous ion, which, on losing an electron, can become ferric ion again. In shuttling between the two forms, the iron atom can transfer one electron at a time. Copper can do the same in shuttling from cuprous ion (Cu^+) to cupric ion (Cu^{++}), and molybdenum also can behave in this fashion.

Such one-electron transfers, possible to the metal-containing flavo-enzymes and impossible to the purely organic py-

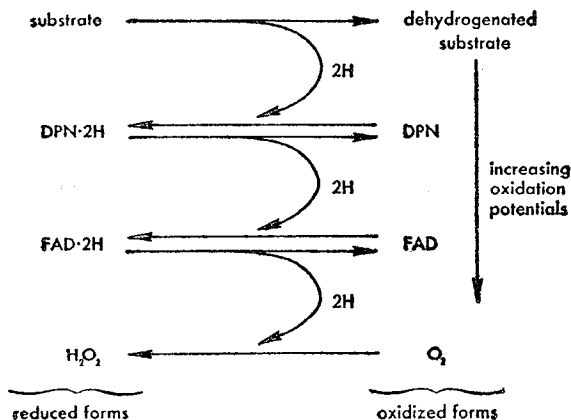


Figure 64. *Dehydrogenation in Steps*

ridino-enzymes, are apparently necessary to the lowering of the energy of activation involved in transferring hydrogen to the oxygen molecule. The details, however, remain unknown as yet.

One might almost suppose that the situation with regard to the hydrogen atoms obtained in the course of catabolism was now satisfactorily straightened out, but of course it is not. When oxygen accepts hydrogen atoms from the flavo-enzymes, hydrogen peroxide is formed. Perhaps some actually is in the body, in this fashion, and catalase must be present to destroy the hydrogen peroxide as it is formed. On the whole, however, hydrogen peroxide is *not* formed and so the scheme as I have presented it so far cannot be complete. Something additional is still required.

In 1925 the British biochemist D. Keilin studied the manner in which suspensions of minced tissues of various sorts, from brain to bacteria, absorbed light. He found that some half dozen bands of absorption occurred, which he attributed to a compound he called "cytochrome" (from Greek words meaning "cell color"). Further study showed

that the bands of absorption existed in pairs, each pair being attributed to a separate substance. These were, naturally enough, named "cytochrome *a*," "cytochrome *b*," and "cytochrome *c*." As time went on, even these divisions proved not to represent single compounds and, in particular, there was a compound closely related to cytochrome *a* that eventually received the name "cytochrome *a₃*."

The only cytochrome that could be easily separated from the tissue suspensions was cytochrome *c*. It turned out to be a relatively small protein, with a molecular weight of 13,000, and with each molecule containing a single iron atom. This iron atom formed a part of a heme group, the same heme that occurs in hemoglobin (see Chapter 18). Further study showed that the other cytochromes also contained iron; thus, in each case, forming part of a group that was either heme or something very like it.

The cytochromes can be lumped together under the name "heme-enzymes," in consequence. Catalase, which I mentioned in the previous chapter, is another example of a heme-enzyme, but one that doesn't share the functions of the cytochromes. Hemoglobin is a heme-protein, but not a heme-enzyme. (Not all enzymes that possess iron as part of the molecules are heme-enzymes. Succinic acid dehydrogenase contains iron, but that iron does not form part of a heme group.)

When the chain of hydrogen carriers (the "respiratory chain") was being worked out, it quickly became clear that the cytochromes must form part of it. They were almost universally found in cells, the only exceptions including certain bacterial cells that are "obligate anaerobes"; that is, cells that cannot live except in the *absence* of oxygen. The very fact that these cells don't have cytochromes and also cannot utilize oxygen is in itself strong evidence in favor of the role of cytochromes in the respiratory chain.

Then, too, any substance that interferes with the workings of the cytochromes—and, specifically, with the shuttling of the iron atom between ferrous ion and ferric ion, so as to stop the electron transfer—ends oxygen absorption. The cyanide group ($\text{—C} \equiv \text{N}$) in such compounds as prussic acid (HCN) and potassium cyanide (KCN) freezes the iron atom in the ferrous form, for instance, and stops respiration quickly. (Which is what makes cyanide so poisonous.)

The position of the cytochromes in the chain can be

judged from the oxidation potentials, which are given in Table 9. Apparently the cytochromes come after the flavins, accepting the electrons of the hydrogen atoms and passing them on, one at a time, by way of ferrous ion/ferric ion shuttle, from *b* to *c* to *a* to *a*₃. Presumably the hydrogen atoms themselves are also passed on, though the details of this part of the transfer are lacking.

Yet the chain must come to an end somewhere. At some point the hydrogens must be handed over to oxygen, and cytochrome *a*₃ is that point. The oxygen/water system has an oxidation potential of +0.80 and it finally takes over. Since cytochrome *a*₃ can use oxygen as a hydrogen acceptor, with the formation of water, *not* hydrogen peroxide (it is the enzyme that finally catalyzes the breaking of the O—O bond), it is an example of an "oxidase." In fact the most common name of cytochrome *a*₃ is "cytochrome oxidase."

And, finally, with the formation of water, we can relax. The respiratory chain is present in full.

But, once again, why so long? Why the number of co-enzymes and prosthetic groups engaged in the transfer of the hydrogen atoms from substrate to oxygen?

Well, the dehydrogenation of a compound represents, as I said earlier in the book, a free energy decrease of anywhere from 35 to 70 kilocalories per mole, and yet this is of no use to the body unless at least some of it is converted into high-energy phosphate bonds. A single reaction, as far as is known, cannot produce more than one high-energy phosphate bond. Since these involve an increase in free energy of only about 8 kilocalories per mole, the conversion of one dehydrogenation to one high-energy phosphate bond would waste as much as seven eighths of the free energy.

Table 9
OXIDATION POTENTIALS

Oxidation-reduction System	Oxidation Potential
DPN/DPN • 2H	— 0.32
FAD/FAD • 2H	— 0.22
<i>b</i> (Fe ⁺⁺⁺)/ <i>b</i> (Fe ⁺⁺)*	— 0.05
<i>c</i> (Fe ⁺⁺⁺)/ <i>c</i> (Fe ⁺⁺)*	+ 0.25
<i>a</i> (Fe ⁺⁺⁺)/ <i>a</i> (Fe ⁺⁺)*	+ 0.29
<i>a</i> ₃ (Fe ⁺⁺⁺)/ <i>a</i> ₃ (Fe ⁺⁺)*	+ 0.30
* cytochromes	

By establishing a long respiratory chain, living tissue man-

ages to break the free energy decreases into fragments, so to speak, and several of these can each serve for conversion into high-energy phosphate bonds. In the middle 1950s it was shown quite convincingly that a single dehydrogenation could be used to produce three high-energy phosphate bonds as a reasonable average. The efficiency is thus at least 35 per cent.

Naturally, one would like to know where, in the respiratory chain, the high-energy phosphate bonds are formed. Evidence for this can be obtained by studying short cuts in the chain.

Thus, a system that contains lactic acid dehydrogenase will produce three high-energy phosphate bonds for every lactic acid molecule consumed. On the other hand, a system that contains succinic acid dehydrogenase will produce only two high-energy phosphate bonds for every succinic acid consumed. Succinic acid dehydrogenase is a flavo-enzyme, so that in the dehydrogenation of succinic acid the pyridino-enzyme step is left out. Since one high-energy phosphate bond is also left out, it seems reasonable to assume that one such bond must ordinarily be produced in the passage from DPN to FAD.

By adding proper chemicals to respiratory systems, it is possible to uncouple the chain at various joints, reducing the number of high-energy phosphate bonds produced still further. As a result, a reasonable suggestion, at the moment, for the sites of production is shown in Fig. 65.

But what are the reactions by which "oxidative phosphorylation" (that is, the production of high-energy phosphate bonds through the oxidation represented by the transfer of hydrogen atoms) takes place? Ah, that is as yet unknown. And if I ever give the impression in this book that all problems are solved, let me disown that strenuously, and use this problem as an example (one of very many) to the contrary.

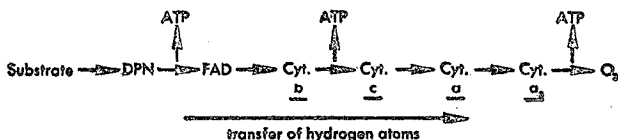


Figure 65. *Oxidative Phosphorylation*

Without knowing the exact details of high-energy phosphate bond production, it is nevertheless possible to count the numbers produced at each stage in the catabolism of lactic acid, and the results are indicated in Fig. 66.

As you see, the catabolism of lactic acid to carbon dioxide and water, which in the calorimeter represents a decrease of 325 kilocalories per mole,¹ gives rise to the production of 18 molecules of ATP. Counting each high-energy phosphate bond formed as representing an increase in free energy of 8 kilocalories per mole, there has been a total storage of 144 kilocalories of chemical energy for each molecule of lactic acid catabolized. Now 144 out of 325 represents an efficiency of just about 45 per cent, which is respectable indeed.

Where in the cell does this oxidative phosphorylation take place? I have not anywhere in this book attempted to describe the cell or to give an idea of its complex structure. And complex it is, for it possesses differentiated regions ("organelles") with differentiated properties, all proving more elaborate as the cytologist's tools grow sharper.

Every complete cell, to begin with, has a small inner portion, the "cell nucleus," marked off from the rest of the cell (the "cytoplasm") by a thin membrane. The primary concern of the nucleus is in cell reproduction and in the accurate transfer of the genes controlling chemical characteristics from mother to daughter cells.² It is not, however, involved in energy production, and within the nucleus there are no flavo-enzymes or cytochromes. Dehydrogenations involve no more than the pyridino-enzymes, which themselves make no use of oxygen as a hydrogen acceptor. The nucleus is, in other words, an anaerobic system and, in searching for the site of the oxygen-consuming energy-producing reactions, we must turn to the cytoplasm.

The various organelles of the cytoplasm (and of the nucleus, too, for that matter) react differently to this dye and that, absorbing one but not another, so that by working with appropriate "stains" it is possible to have one particular organelle absorb color while others do not. In

¹ In Chapter 20, I used the figure 650 kilocalories and if the memory of that is a source of confusion to you, look back and you will see that the equation for which the higher figure was used involved *two* molecules of lactic acid.

² For details on that, I refer you again to my book *The Wellsprings of Life*.

this way, ordinarily undistinguished and difficult-to-see cell particles are made to stand out in garish hues.

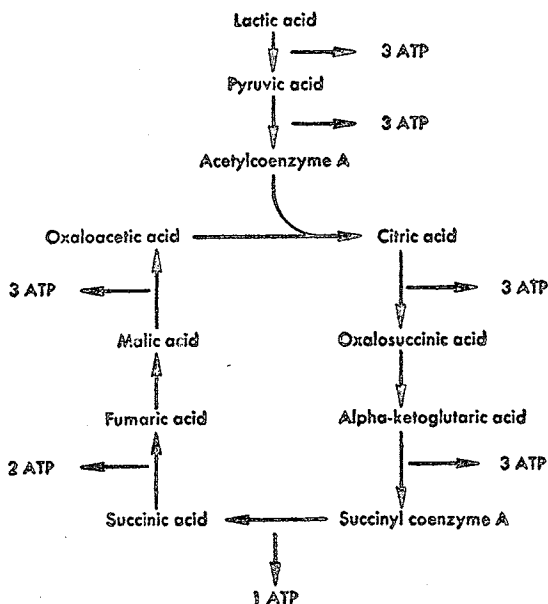


Figure 66. *ATP and the Krebs Cycle*

In 1898 a German cytologist, C. Benda, using a complex stain, found that small granules in the cytoplasm of a wide variety of cells were colored into conspicuous contrast. He called the granules "mitochondria" from Greek words meaning "cartilage-threads," which is a poor name because they are not cartilage. Nevertheless, the name, like many another misnomer, persists and will continue to do so. The singular of the word, following Greek usage, is "mitochondrion."

For a generation and more, mitochondria were mere particles in the cell, their function unknown. In the 1930s, however, with the development of ultracentrifugal methods (see Chapter 16), it became possible to break up cells in various ways and then to spin them down so that different portions

would settle at different rates. In this way, suspensions of pure mitochondria were eventually obtained.

Through the 1940s, moreover, a new kind of microscope of unprecedented powers of magnification was developed. Instead of using light rays focused by lenses, beams of electrons focused by magnetic fields were used. Such "electron microscopes" were made to order for the study of the structure of mitochondria, which are bodies only one to three microns in diameter, and are therefore only about a thousandth the size of the cell itself.

The electron microscope revealed the mitochondrion to be bounded by a double membrane that bellies into the body of the organelle, dividing it into segments and converting it into something that is almost all surface membrane.

Suspensions of mitochondria were found to catalyze all the reactions of the Krebs cycle. It is clear, therefore, that the mitochondria are the "powerhouses" of the cell; that their membranes are actually conglomerations of all the enzymes and coenzymes needed for catabolizing foodstuffs and producing high-energy phosphate bonds. It has been estimated that an individual mitochondrion may contain as many as 10,000 separate assemblies; each capable of conducting a complete catabolism; each ripping off hydrogen atoms and producing high-energy phosphate bonds.

Detailed knowledge concerning the molecular setup within the mitochondrion awaits deeper probing into these tiny bits of the cell. Scientific instruments are already reaching down to molecular levels, however, and the future looks bright.

Chapter 24

THE JOINING OF THE ROUTES

The last four chapters have been given over, in one way or another, to the catabolism of glucose, first to lactic acid by anaerobic glycolysis, then to carbon dioxide and water by way of the Krebs cycle. This is not to say that

these are the only routes by which the body may form high-energy bonds and store chemical energy.

For instance, in the early 1950s the British biochemist F. Dickens, and others, were able to show that in a variety of plants and microorganisms, and in some animal tissues as well, glucose could be catabolized to the three-carbon stage by a route other than anaerobic glycolysis. This new route, which joins the usual pathway about halfway down the line of ordinary glycolysis, is sometimes called the "Dickens shunt." It goes by way of five-carbon and four-carbon sugars, and is not anaerobic. It makes use of the respiratory chain and atmospheric oxygen.

The value of an alternate route is twofold. First, if something should happen to inhibit the working of one route temporarily, the organism is not entirely without resources. It can limp along with the alternate route.

Secondly, any catabolic route supplies not only energy but also a number of metabolic intermediates. These intermediates can serve the body as building blocks for a variety of anabolic reactions. The various components of the Krebs cycle, for instance, can serve as the basis for the formation of some of the amino acids. Oxaloacetic acid, by the addition of an amine group, becomes aspartic acid; alpha-ketoglutaric acid, by the same treatment, becomes glutamic acid (which is why those two amino acids, for instance, are not dietarily essential).

There is thus an advantage in the multiplication (within reason) of metabolic routes, since the body is in this way supplied with a greater number of possible building blocks. The Dickens shunt provides the organism with several five-carbon sugars and related compounds. These can be readily used for the formation of the all-important giant molecules of the nucleic acids, which contain five-carbon sugars as an integral portion of their structure.

Anaerobic glycolysis and the Krebs cycle remain the major routes of glucose catabolism, however, and there would be no profit in going into detail as far as the alternate routes are concerned.

However, in another respect that we cannot so easily ignore, the chapters on catabolism are incomplete. So far I have only talked about the catabolism of glucose. But in Chapter 19, I pointed out that lipids and proteins, particularly the former, are important sources of energy also. How

are they handled? Surely not in the same manner as the carbohydrates.

Well, let's see.

In the first decade of the twentieth century, when Harden and Young were initiating the study of the intermediary catabolism of glucose, the German biochemist Franz Knoop did the same for the intermediary catabolism of lipids.

Knoop used the technique of feeding dogs certain food-stuffs and then studying the material excreted in the urine. In doing so, however, he added a vital refinement of the greatest importance. Had Knoop just fed the dog lipids, the dog would simply have produced water and carbon dioxide and there would have been no way of telling which molecule of water or carbon dioxide comes from the lipid and which from anything else. There had to be some way of putting a label or tag on the lipid molecule so that its remnant could be identified after the dog was through with it. For this purpose, Knoop fed the dog "phenyl-derivatives" of various fatty acids.

Let me explain what that means. To begin with, the fatty acids are the chief products of lipid digestion. Their molecules consist of long hydrocarbon chains, usually made up of sixteen or eighteen carbon atoms, of which the carbon atom at one end is part of a carboxyl group. A common eighteen-carbon fatty acid is "stearic acid." If such a compound is treated in such a way as to attach a benzene ring to the chain at the end opposite to the carboxyl group, the result is a phenyl-derivative, as shown in Fig. 67.

A phenyl-derivative of a fatty acid presents the organism with something it can't readily handle. There are no enzymes in a dog (or a man) capable of breaking up a benzene ring or splitting it off a carbon chain. The benzene ring is on the fatty acid pretty much for keeps. The dog's tissues may succeed in catabolizing the rest of the fatty acid but in the end it can only get rid of the benzene ring (with any fragment of fatty acid that may still be hooked to it) by excretion in the urine. By searching for any compound with a benzene ring in the urine, Knoop could feel reasonably sure he would, if successful, find a remnant of the fatty acids he had fed the dog.

Thus, the fatty acids are tagged. Knoop had made use of a "tracer" (that is, something by which the fate of a

compound during metabolism could be traced). Before describing the results Knoop obtained in these experiments,

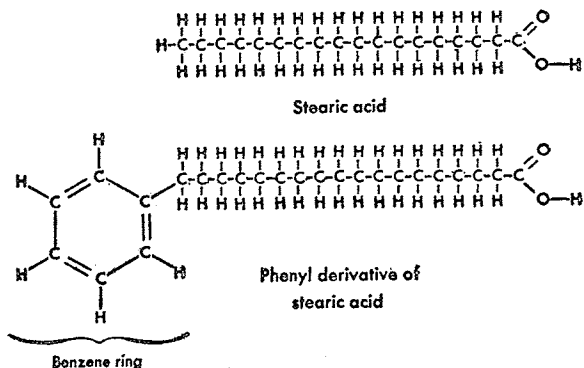


Figure 67. Phenyl Derivative of Fatty Acids

however, I would like to follow the subsequent history of tracers, for they turned out to be an extremely important technique in biochemistry.

The immediate objection to Knoop's technique was that he did, in fact, introduce compounds into the body that it was not equipped to handle. They were "unphysiologic." Even if they were not toxic and if the organism continued in perfect health as far as could be told, how could anyone be sure that the organism's catabolic routes were not affected by the presence of the tag? Perhaps the benzene ring inhibited some important enzyme. In short, how could one be certain that the metabolism that took place in the presence of the tag would be precisely like that which would take place in its absence?

The ideal situation would be to use a tagged compound that the body could not differentiate from the normal untagged compound, but which the experimenter *could*. This might seem to be asking too much of the universe, but, as it happened (by one of the happiest strokes of fortune in the history of science), within a decade of Knoop's experiments, the existence of "isotopes" was recognized.

It was shown in 1913, by the British physicist Frederick Soddy, that the atoms of a particular element were not

necessarily identical after all. All had the same number of protons in the nucleus, to be sure, and therefore the same number of electrons in the outskirts, and therefore the same chemical properties (see Chapter 17). However, different atoms of the same element might have different numbers of neutrons in the nucleus, so that the resulting atoms would differ in atomic weight. They would be isotopes of that element.

For instance, some carbon atoms would have nuclei containing six protons and six neutrons, for an atomic weight of 12, while others would have six protons and seven neutrons for an atomic weight of 13. Symbolically, these isotopes would be differentiated as C^{12} and C^{13} . Each would have six electrons, two in an inner shell and four in the outer; each would therefore engage in the same chemical reactions within the body. The body would not distinguish between compounds containing either.

Then, in 1919, the British physicist Francis William Aston developed the "mass spectrograph," a device that could be made to bend speeding ions in a magnetic field. The heavier the ions, the more gently they are bent. If some ions contained a C^{12} atom and others, otherwise identical, contained a C^{13} atom, the two varieties could be separated. Each in its differently curving path would hit a different spot on a photographic plate and the quantity of each could be measured by the extent of the blackening of the film. In this way, for instance, it could be shown that ordinary carbon was made up of C^{12} and C^{13} in the ratio of 989 to 11.

Suppose, now, that a compound could be prepared with all its carbon atoms in the form of C^{13} . To the body this would make little difference. It would be treated as would any other compound of the same type with C^{12} as the carbon atom. The C^{13} substance might be converted into a second compound that would then contain an abnormally high quantity of C^{13} . Any compound that contained that abnormally high quantity would have to be derived from the original isotope-tagged substance. By burning various compounds to carbon dioxide and running the gas through the mass spectrograph under appropriate conditions, the experimenter can detect any abnormal quantities of C^{13} that might be present. From its presence or absence, he can initiate a chain of reasoning that would illuminate the metabolic route along which the compound had passed.

Though this may sound easy in the telling, it proved extraordinarily difficult in the doing. Relatively rare isotopes such as C^{13} must be concentrated, compounds formed out of it, other compounds obtained in reasonable purity from the organism, and in the end the rather difficult techniques of mass spectrography had to be applied. As a matter of fact, such were the difficulties that the use of isotopic tracers remained of negligible importance until 1932 when the American chemist Harold C. Urey discovered "deuterium" or "heavy hydrogen."

Where ordinary hydrogen has a nucleus consisting of nothing more than a proton, so that it has an atomic weight of 1, deuterium has a nucleus made up of a proton and a neutron, so that it has an atomic weight of 2. The symbols are H^1 and H^2 respectively. Whereas the difference in weight between C^{12} and C^{13} is one in twelve or 8 per cent, that between H^1 and H^2 is 100 per cent.

Advantage can be taken of this unusual percentage difference in mass between ordinary hydrogen and deuterium. Heavier isotopes react just as lighter ones do but more sluggishly. For elements other than hydrogen, the difference is negligible; for hydrogen the difference in mass is so great that deuterium is noticeably sluggish. Water that contains deuterium in its molecule ("heavy water") distills more slowly and electrolyzes more slowly also. If large quantities of water are distilled or electrolyzed, the final fraction is high in heavy water. From such heavy water, organic compounds high in deuterium can be prepared. Furthermore, deuterium-containing end products can be detected and analyzed with comparative ease by the mass spectrograph.

In 1935 the German-born American chemist Rudolf Schoenheimer used deuterium to tag fatty acids, much as Knoop, a generation earlier, had used the benzene ring. Later, with the techniques further developed, Schoenheimer used "heavy nitrogen," N^{15} (ordinary nitrogen atoms being N^{14}), to tag amino acids. As a result of his experiments, he was able to show that the body constituents are extraordinarily active. Even where there is no over-all change visible to the chemist's probing eye, there are constant shiftings and transferrals, joinings and breakups, shufflings and reshufflings of atoms.

But, even so, isotopic tracer technique remained clumsy. Once an isotope-tagged compound was added to the body,

it was diluted through all the body's substance and the subsequent isolation and detection of the rare isotope remained a delicate and time-consuming job.

Again, there proved a way out. Some isotopes are made up of atoms with nuclei containing proton-neutron combinations that are not stable. Various subatomic particles are hurled out of such nuclei at fixed and calculable rates. For instance, C^{14} , with a nucleus made up of six protons and eight neutrons, hurls energetic electrons outward at such a rate that half of any quantity of C^{14} is changed to N^{14} (the most common nitrogen isotope) in a little over five thousand years.

Such exploding nuclei are "radioactive." If the radioactive C^{14} is contained in compounds taken into the body, it will behave like ordinary stable C^{12} or C^{13} as far as metabolism is concerned—but it will leave a trail of energetic particles that the experimenter can follow with relative ease. The C^{14} can be traced with precision even when it is present in vanishingly small quantities.

It remained only to find ways of manufacturing radioactive isotopes in quantity. Prior to World War II, only the naturally radioactive atoms were available to any extent and these, unfortunately, were for the most part isotopes of elements that did not occur in the body or take part in metabolism.

Then, during World War II, the nuclear reactor was developed and radioactive isotopes, not ordinarily present in nature, could be manufactured in relatively large quantity. Since World War II, virtually all studies of metabolism have involved the use of compounds tagged with C^{14} , H^3 , Na^{24} , and other such radioactive isotopes. Life without them would now be just about impossible for biochemists.

But to get back to Knoop, and to his use of the first tagged compounds. . . .

When Knoop fed dogs tagged fatty acids containing an even number of carbon atoms, he obtained from the urine a benzene ring to which a two-carbon group was attached; a compound named "phenylacetic acid." This, in turn, was attached to a molecule of glycine. When he started with tagged fatty acids containing an odd number of carbon atoms, he ended with a benzene ring to which a single carboxyl group was attached ("benzoic acid") and this, again, was attached to a molecule of glycine (see Fig. 68). The

addition of glycine is a routine method used by an organism to make an unwanted molecule more easily removable

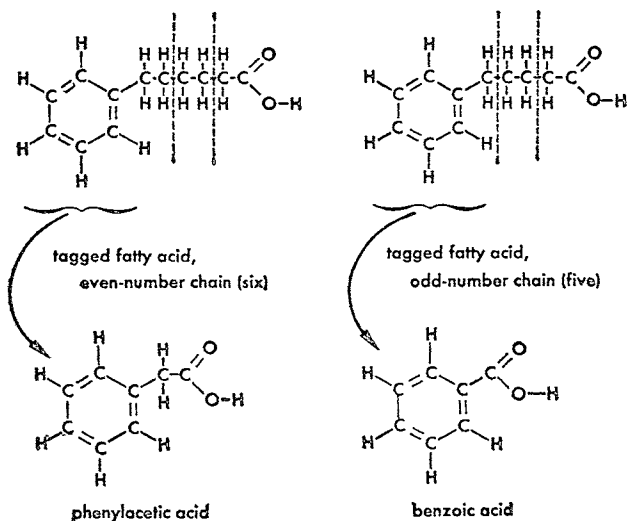


Figure 68. *Fatty Acid Catabolism*

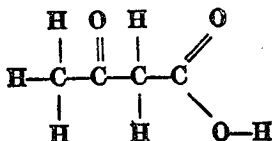
through the kidney, so Knoop ignored that part of the molecule and considered the phenylacetic acid and the benzoic acid.

Knoop explained the results by suggesting that the long chain of the fatty acid was cut up two carbon atoms at a time, starting at the end with the carboxyl group. When the total number of carbon atoms was even, the last bit consisted of two carbon atoms attached to the benzene ring (a connection the body could not break). When the total number was odd, the last bit consisted of one carbon atom attached to the benzene ring.

It also seemed logical that fatty acids were built up in increments of two-carbon groups. Beginning with one such group and adding others of the same sort, only even-number chains of carbon atoms can be obtained and, as it happens, the fatty acids making up the lipids of all forms of life almost invariably contain even numbers of carbon atoms. That

is, eighteen-carbon fatty acids and sixteen-carbon fatty acids are extremely common in nature, while a seventeen-carbon fatty acid is virtually unheard of. (The fatty acids with odd numbers of carbon atoms that Knoop fed to his dogs were synthetics, products of the laboratory.)

Two years later the German biochemist Gustav Embden added another refinement. He found that when liver tissue was incubated with fatty acid a four carbon remnant was located in solution. This was "acetoacetic acid."



This and several related substances are called "ketone bodies" because of the $\text{C} = \text{O}$ ("keto group") combination in acetoacetic acid. Acetoacetic acid is an example of a fatty acid in which a two-carbon piece is in the process of being cut off, as I shall explain shortly, and this strengthened Knoop's two-atoms-at-a-time hypothesis.

The ketone bodies do not occur in perceptible quantities in the blood and urine of intact animals, even if isolated liver does produce them. However, ketone bodies do occur in the blood and urine of diabetics. Since diabetes is primarily a disorder of carbohydrate metabolism, the fact that it has such a pronounced effect on ketone bodies, which are products of lipid metabolism, shows the interconnection between the two metabolic routes.

Despite this early start, the details of fatty acid catabolism remained obscure until the days of isotope-tagging dawned. It was not until 1951, in fact, that biochemists could write the equations of the changes involved with any degree of confidence at all.

Current knowledge confirms Knoop's original surmise: that the fatty acid carbon chain is snipped to pieces two carbons at a time, starting at the carboxyl end. Two refinements are added. First, the fatty acid condenses with Coenzyme A before it can undergo catabolism. Secondly, the cutting off of two carbons takes place in four stages (see Fig. 69).

The first stage is a dehydrogenation, a hydrogen atom being removed from each of the two carbons adjacent to the

carboxyl group. The second stage is a hydration, the elements of water being added so that a hydroxyl group is

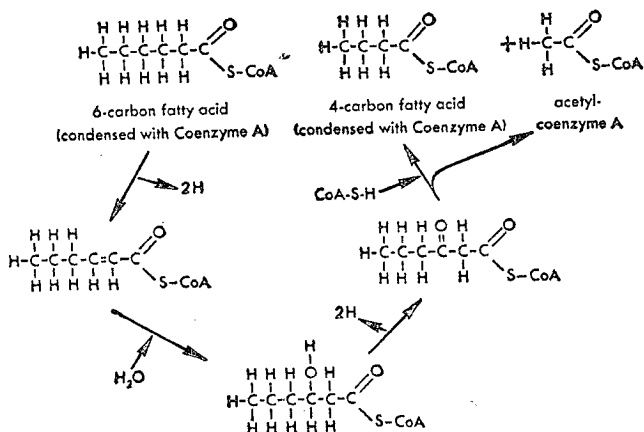


Figure 69. *Fatty Acid Oxidation Cycle*

added to the second carbon from the carboxyl group (this second carbon is the "beta carbon" and the process is therefore called "beta oxidation"). In the third stage, two more hydrogen atoms are removed so that the beta carbon becomes part of a keto group. If you look back at the formula of acetoacetic acid, you will see that this must be the product of this third stage. In the fourth stage, the two end carbons are split off altogether as acetylcoenzyme A, and what is left of the chain is a new (but two carbons shorter) fatty acid, which condenses with Coenzyme A and goes through the changes again.

This is the "fatty acid oxidation cycle." Actually the process is a helix rather than a cycle because it doesn't quite return to its starting point, but to a fatty acid two carbons shorter than the original, then to a fatty acid two carbons shorter still, and so on. In this way an eighteen-carbon fatty acid, such as stearic acid, is eventually converted to nine molecules of acetylcoenzyme A.

Acetylcoenzyme A, once produced, will combine with oxaloacetic acid to form citric acid and thus enter the Krebs cycle (as described in the previous chapter). It doesn't mat-

ter whether acetylcoenzyme A is formed from fatty acid catabolism or from glucose catabolism; the body has no way of telling its past history (see Fig. 70).

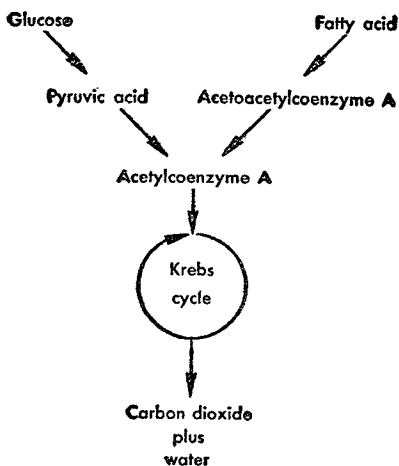


Figure 70. *The Joining of Carbohydrate and Lipid*

Acetylcoenzyme A can also be recombined to form, eventually, fatty acids, by reversing the fatty acid oxidation cycle. That is, first two molecules of acetylcoenzyme A combine to form acetoacetylcoenzyme A, then two hydrogen atoms are added, water is removed, two hydrogens are added again—and a four-carbon fatty acid, condensed with Coenzyme A, is the result. Continue this by adding another acetylcoenzyme A, and eventually the chain is as long as you like.

This means that acetylcoenzyme A is nothing more nor less than the “common building block” whose presence I deduced in Chapter 19 as being the logical way of explaining the fact that starchy foods are fattening and that blood glucose can be maintained at normal level by use of the fat reserves of the body.

The respiratory chain is, of course, involved in fatty acid catabolism, as it is in glucose catabolism. Each turn of the

fatty acid oxidation cycle includes two dehydrogenations. The first is catalyzed by a flavo-enzyme and the second by a pyridino-enzyme. We can assume the formation of two molecules of ATP in the first case and three in the second, so that each turn of the cycle produces five molecules of ATP by oxidative phosphorylation.

For a fatty acid such as the eighteen-carbon stearic acid, eight turns of the cycle are required to produce nine molecules of acetylcoenzyme A (*not* nine turns; cut an 18-inch strip of paper into 2-inch pieces and you will see that you need eight cuts and not nine to perform the task).

Thus the conversion of stearic acid to nine molecules of acetylcoenzyme A produces 8 times 5, or 40 molecules of ATP. If you check back to the diagram of the Krebs cycle on p. 330, you will see that from acetylcoenzyme A onward through that cycle, 12 molecules of ATP are formed. Nine molecules of acetylcoenzyme A will therefore form 9 times 12, or 108 molecules of ATP.

Thus, in passing from stearic acid to carbon dioxide and water, no less than 40 plus 108, or 148, molecules of ATP are formed. Nor is this all. A typical molecule of a lipid, such as "stearin," consists of a three-carbon alcohol, glycerol, condensed with three molecules of stearic acid. Each of the three stearic acid molecules will form 148 molecules of ATP, for a total of 3 times 148 or 444 molecules of ATP. The glycerol can be converted to pyruvic acid, which is good for 15 molecules of ATP itself, making a grand total of 459 molecules of ATP formed for each molecule of stearin completely catabolized.

Now a gram of lipid liberates 9 kilocalories of heat on conversion to carbon dioxide and water (see Chapter 14). A mole of stearin weighs 891 grams, so that the heat produced by conversion of a mole of stearin to carbon dioxide and water is 891 times 9, or 8019 kilocalories. In forming 459 molecules of ATP, 459 times 8, or 3672 kilocalories of chemical energy are stored. The efficiency is 3672/8019 or just over 45 per cent. As such things go, this is excellent.

But if you will look at the figure on p. 330, you will see that acetylcoenzyme A enters the Krebs cycle by combination with oxaloacetic acid. Oxaloacetic acid, however, is produced in quantity only by carbohydrate metabolism (and to some extent by protein metabolism) but *not* by lipid metabolism.

If carbohydrate is lacking in the diet, while lipid is plentiful, much acetylcoenzyme A is formed but little oxaloacetic acid. As acetylcoenzyme A accumulates, it condenses with itself to form acetoacetic acid and the ketone bodies begin to accumulate, bringing about a condition known as "ketosis."

A high-lipid, low-carbohydrate diet is "ketogenic" because it favors the formation of ketone bodies. Fasting is also ketogenic because after the first day the carbohydrate stores of the body are gone and energy must be obtained from lipid mainly. When liver is incubated with fatty acid, but with no added carbohydrate, again there is ketosis, as Embden discovered.

No great harm is done by fasting or by ketogenic diets, because neither is particularly attractive and the average man does not follow either course voluntarily. Then, even if forced into ketosis in this way, a small quantity of ingested carbohydrate will produce an "anti-ketogenic" effect and will right matters.

It is in diabetes that ketosis is most serious. If the disease is untreated, there is a permanent disorder of carbohydrate metabolism that produces an increasingly severe shortage of oxaloacetic acid and therefore a continually intensifying ketosis. It is the toxic effect of the ketone bodies that is usually the cause of the final coma and death of the diabetic.

(Not all lipids, to be sure, have as their prime function the production of energy. The lipid molecules I have been discussing in this chapter are "simple lipids," constructed of glycerol and fatty acids. There are, on the other hand, "compound lipids" which, in addition, contain phosphate groups and any of several nitrogen-containing groups. There are also compounds such as "cholesterol" with the properties of lipids but with an entirely different molecular structure.

(These latter compounds have specialized functions in the body. The phosphate-containing lipids make up the mitochondrial membrane in which—as I pointed out in the previous chapter—the various enzymes of the Krebs cycle and the respiratory chain are embedded. As for cholesterol, that makes up a major portion of the fatty sheath of the nerves.

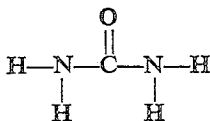
(The compound lipids and cholesterol contain as large a store of energy as do the simple lipids, but the role of the former in the cell machinery is so vital that their energy stores are not drawn upon even in the last extremes of starvation. It is only because they play no part in energy

production that I neglect these supremely important compounds in this book.)

Now let us turn to the proteins, which, after digestion, enter the body in the form of amino acids. Before these can be utilized for the production of useful energy they must be stripped of their nitrogen.

In 1773 the French chemist G. F. Rouelle (Lavoisier's teacher) discovered a nitrogenous compound in urine and named it "urea" after its source. Once the composition of proteins began to be studied at the beginning of the nineteenth century, urea was at once recognized as the obvious route by which the body excreted the nitrogen of protein.

Its formula was shown to be



or, more briefly, NH_2CONH_2 , once structural formulas became the order of the day. As it happens, urea was involved in two startling advances in biochemistry. It was the first organic compound to be synthesized from an inorganic starting material (see Chapter 13) and the enzyme catalyzing its breakdown was the first to be crystallized (see Chapter 15).

The method by which the body produced urea remained unknown throughout the nineteenth century. In the first decade of the twentieth century, however, when Harden and Young on one hand and Knoop on the other were making a beginning with respect to the intermediary metabolism of carbohydrates and lipids, the German biochemist Albrecht Kossel and his English coworker Henry Drysdale Dakin were doing the same for the intermediary metabolism of protein.

They discovered an enzyme called "arginase" that *catalyzed* the hydrolysis of the amino acid, arginine. In so doing, urea is produced by means of the reaction shown in Fig. 71. The truncated molecule of arginine left behind is the amino acid "ornithine," which occurs only as an intermediary in urea formation and does not occur as part of the various protein molecules.

The importance of arginase was emphasized in several ways. Urea is produced by the liver (for if the liver of an

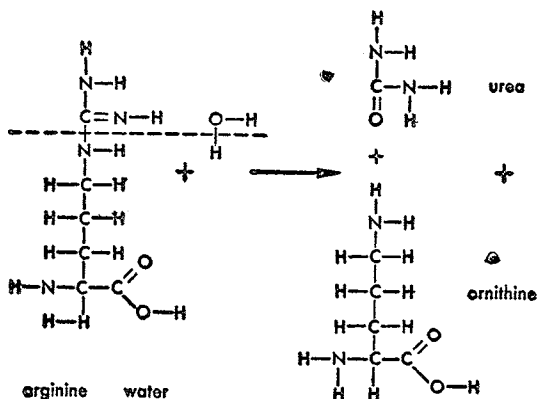


Figure 71. *The Hydrolysis of Arginine*

animal is removed, urea production stops) and arginase is most abundant in the liver. Those animals, such as mammals and amphibia, which produce urea, possess arginase. Animals, such as the bird and reptiles, which produce other nitrogenous end products, do not possess arginase. Clearly, arginine and arginase are involved in whatever mechanism is used by the body to form urea.

But a man may easily produce ten to thirty-five grams of urea in a day without depleting his body of its arginine supply. Arginine must be reconstituted from ornithine.

This was made clear in 1932 by Krebs (who was later to work out the Krebs cycle) and his coworker, K. Henseleit, using liver slices. They found that the addition of arginine stimulated the production of urea, which was expected. So did the addition of ornithine, and that strengthened the possibility that liver converted it into arginine. Finally, they found out that "citrulline," a compound resembling arginine in structure, also stimulated urea-production.

To explain all this, Krebs and Henseleit evolved a scheme usually called the "urea cycle," which is pictured in Fig. 72. As you see, in this scheme ornithine, citrulline, and arginine are continually being regenerated. However, a molecule of

ammonia and one of carbon dioxide are added to the system in converting ornithine to citrulline, and a second molecule

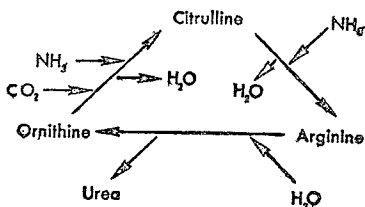
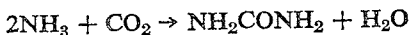


Figure 72. *The Urea Cycle*

of ammonia is added in converting citrulline to arginine. As for water, one molecule is added to the system in the hydrolysis of arginine but a molecule of water is subtracted from the system in each of the other two steps, so that the net effect is the subtraction of one molecule of water. And, finally, a molecule of urea is subtracted when arginine is hydrolyzed. The over-all reaction, then, is:



Of course, this version of the urea cycle is a simplification. No one really thought, at the time it was proposed, that ammonia molecules, as such, were added. Ammonia is highly toxic and there is no ammonia to speak of anywhere in the body at any time. It was not until the 1950s, however, when isotope-tagging experiments could be conducted, that the details of the conversions were worked out.

The nitrogen-containing groups are added, as it turns out, in the form of amine groups obtained from amino acids. In going from ornithine to citrulline, glutamic acid (one of the commonest of the amino acids) donates both an amine group and the equivalent of a carbon dioxide. In doing so it must make use of a high-energy phosphate bond obtained from ATP. In going from citrulline to arginine, another amino acid, aspartic acid, donates an amine group, again at the expense of ATP.

Thus, the production of urea is an energy-consuming reaction. Each turn of the urea cycle consumes two molecules of ATP.

Nor need we think that glutamic acid and aspartic acid are the only amino acids to contribute their amine groups for the formation of urea. In the 1930s, Schoenheimer, in his work with N^{15} to which I referred earlier in the chapter, showed that if you began with one amino acid tagged with N^{15} all amino acids (with the exception of lysine, which is a special case) soon carried it. The implication was that the amine group was readily transferred from compound to compound. Enzymes catalyzing such transfers ("transaminases") were discovered and the details worked out. After transamination, a double bound oxygen ($=O$) replaces the amine group ($-NH_2$) (see Fig. 73).

(The transaminases are examples of enzymes that have "pyridoxal phosphate" as the coenzyme and this includes an atom grouping that the body cannot synthesize. Thus, still another B vitamin, "pyridoxine," is accounted for.)

The flow of amine groups, then, is from the amino acids generally to glutamic acid and aspartic acid and from these into the urea cycle. Those amino acids that contain additional nitrogen atoms, or even sulfur atoms, dispose of these by routes I shall not go into. Suffice it to say that in one way or another all nitrogen and sulfur atoms are removed and a "carbon skeleton" remains behind.

This carbon skeleton may be one of the compounds we have already come across in the description of the catabolism of carbohydrate or lipid. For instance, when alanine undergoes transamination, it loses its amine group and gains a doubly bound oxygen, forming pyruvic acid.

The pyruvic acid is decarboxylated, with the aid of Coenzyme A in the usual manner, to form acetylcoenzyme A, which can enter the Krebs cycle. It doesn't matter to the body whether the pyruvic acid arises from glucose or from alanine. Pyruvic acid is pyruvic acid and just as energy-filled either way. Other amino acids, such as valine, serine, threonine, and cysteine, are also converted, somewhat more indirectly, to pyruvic acid.

A few amino acids, leucine as an example, undergo changes that convert them into acetoacetic acid and progress to acetylcoenzyme A and into the Krebs cycle by the lipid catabolism route.

Still other amino acids enter the Krebs cycle directly. For instance, aspartic acid on deamination becomes oxaloacetic acid and glutamic acid on deamination becomes alpha-

ketoglutaric acid, both of which are intermediates of the Krebs cycle and can instantly form a part of it.

Fig. 74 summarizes the catabolic interconnections of proteins, lipids, and carbohydrates.

Of course, by putting various catabolic reactions into reverse, it is possible to produce amino acids out of the appropriate carbon skeletons. For instance, if the body has an oversupply of glutamic acid and an undersupply of alanine, it can shift an amine group from glutamic acid to pyruvic acid and form alanine out of the latter. Or, if the reverse is true, it can shift an amine group from alanine to alpha-ketoglutaric acid and form glutamic acid out of the latter. Amino acids that can be formed by easy reversal of reactions are not essential in the diet.

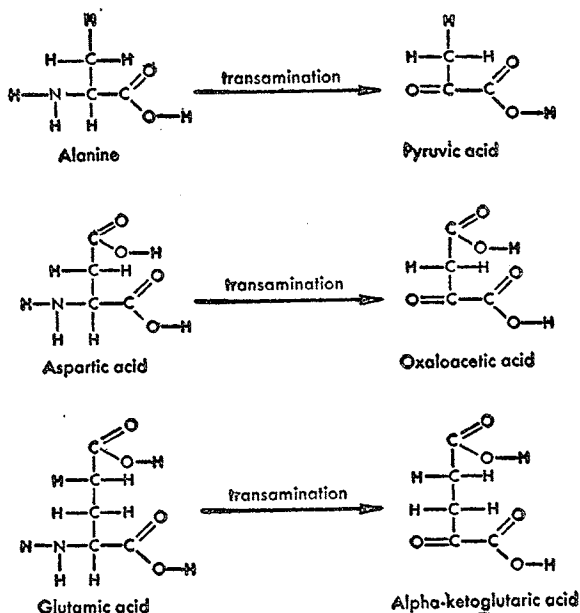


Figure 73. Transamination

On the other hand, some amino acids, during the course of those reactions that produce the carbon skeleton, go

through at least one step that turns out to be irreversible. Such amino acids cannot be formed again from the various metabolic intermediates. They must be present in the diet. These are essential amino acids, which I listed and discussed in Chapter 19.

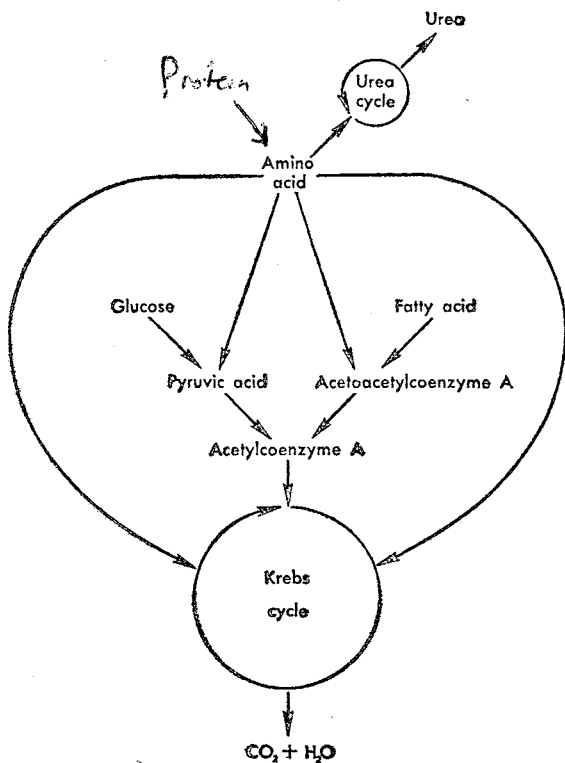


Figure 74. *The Joining of the Protein*

Chapter 25

THE ULTIMATE SOURCE

In the previous two chapters I pointed out that the energy produced by the catabolism of carbohydrates and lipids was stored as high-energy phosphate bonds, with an efficiency of 45 per cent. Though this is a high efficiency, far higher than that of many very useful man-made appliances, there is no hiding the fact that more than half the energy supplied by foodstuffs is lost. (This is not to say it is entirely wasted. Even as raw heat, it helps maintain the body temperature of a warm-blooded creature such as man.)

Furthermore, additional energy is lost when ATP is utilized. Thus, if ATP is used to form a sugar phosphate or a peptide bond, 8 kilocalories are expended in order that a reaction requiring no more than 4 and perhaps as little as 0.5 might go forward.

On the whole, it is estimated that only some 10 per cent of the energy content of food is put to work in some entropy-decreasing fashion. The non-utilization of the remaining 90 per cent contributes to the over-all increase in entropy that is inevitable even for living tissue according to the second law of thermodynamics.

All this implies that ten pounds of digestible food is equivalent to one pound of living tissue. Suppose, then, that one species of organism (the "consumer") obtains its food by eating another species of organism (the "producer"). If the consumer is to maintain its total tissue-mass over the generations, then the producer must maintain its own total tissue-mass at a level at least ten times higher than that of the consumer. The producer, after all, must not only maintain itself; it must also submit to having to support the consumer, with 90 per cent of what it contributes to that support being wasted.

If, in turn, the consumer must serve as food for still another type of organism ("consumer-two"), the consumer will have to maintain a total mass at least ten times that of consumer-two—and so through consumer-three, consumer-four as far as the list extends.

To make this plain by actual example: If lions live on zebras and if both species are maintaining a stable population, then there must be at least ten pounds of zebra in existence for every pound of lion. Then, since the zebras feed on grass, there must be at least ten pounds of grass for every pound of zebra, and therefore at least one hundred pounds of grass for every pound of lion.

This is an example of a "food chain" and it is invariably a sharply dwindling pyramid. There cannot be very many steps in this chain, for less than 10 per cent of less than 10 per cent of less than 10 per cent, and so on, soon represents a very small mass of tissue, even when the mass of the original step is large. The few species that are at the peak of the food chain pyramid—that are large and strong enough to eat fairly large carnivores and too large and strong to be eaten in their turn—may be in an enviable position—but they will never be numerous. The polar bear that lives on seals, or the sperm whale that lives on giant squid, or the killer whale that lives on other whales are expensive and wasteful forms of life that the earth simply cannot support in large numbers. *Wm. B. Davis*

A species may improve its position, as far as the total mass of itself which the earth can support is concerned, by cutting out one or more links in the food chain. If a lion could learn to live on grass instead of on zebras, its food supply would increase at least tenfold. There could then be ten times as many lions in the world, all living as well as lions do now, or each lion could grow to ten times its present mass without straining the food supply.

Of course, lions can't do this, but it is no accident that the largest land animals, the elephant, hippopotamus, and rhinoceros, are herbivores.

To be sure, there were carnivores in the past eras that were far larger than the elephant. Tyrannosaurus, the largest land carnivore that ever lived, had a mass at least twice that of an elephant. But the herbivores of the time, on which it preyed, were larger than modern herbivores. Fur-

thermore, the largest of the giant reptiles of the Mesozoic, such as the brachiosaurus and the brontosaurus, were herbivores.

The ocean, which is far richer in life than the land, supports the largest carnivores of all. The sperm whale is as massive as the largest dinosaurs and is a carnivore, using for food other carnivores (the giant squid) not very much smaller than itself. However, the very largest whale, notably the sulfur-bottom, which attains a mass of up to 150 tons and is the largest animal ever to have existed on earth, past or present, cuts out several links of food chain by feeding on tiny shrimp-like crustacea and other living debris which it filters out of tons of water by the shreds of whalebone that fill its giant maw.

All this applies to man. Not all man's ingenuity can defeat the laws of thermodynamics. True, man can increase his own food supply at the expense of species of organisms he cannot (or will not) eat. That is, he can clear forests and plant grain in their place, killing all unwanted plant life that may try to compete and all insects that may want a place at the table (or at least killing all he can). Again, he may breed animals for food, displacing wild species and killing predators. Still, in the end, man's food supply must have at least ten times the mass of all humanity.

Man, too, may be forced by circumstances to short-circuit the food chain. If man lives on herbivores that feed on vegetation, then there must be at least a hundred pounds of vegetation for every pound of man. If man lives on vegetation directly, perhaps only ten pounds of vegetation must be present for every pound of man. As the population pressure increases, and as the diet of modern Asians already indicates, man finds himself becoming a vegetarian whether he likes it or not.

If the food chain were restricted to living organisms only, life would soon come to an end. The terrible waste of tissue at every meal could not long be supported. The fact that life, nevertheless, is supported in quantity and has been so supported for at least a billion years can only mean that somewhere energy must enter the food chain from a source outside living organisms altogether, and in quantity sufficient to replace the waste.

The possible sources of large-scale energy in the inanimate

environment about us are (1) solar radiation, (2) wind, (3) running water, (4) tides, (5) earth's internal heat, and (6) nuclear reactions.

Of these, the energy of the tides is restricted to a negligible portion of the earth's surface and that of water power to a portion almost as negligible. Wind power, while not so restricted, is impossibly erratic. The earth's internal heat is either negligible or makes itself felt catastrophically in earthquakes and volcanoes. As for nuclear energy (in the form of natural radioactivity, cosmic radiation, and so on), that is very thinly distributed. None of these can be viewed as sufficiently intense, sufficiently gentle, sufficiently widespread, and sufficiently constant to serve as a reliable energy foundation for all of life.

By elimination that leaves us with solar radiation as the only source.

Nor need we be satisfied with negative evidence. The most superficial observation of the world about us must convince us that solar radiation is the ultimate source of life's energy. The food chain ends with green vegetation of one sort or another. It is the "producer." All animal life eats green plants, if the chain is followed down far enough on both land and sea. In the sea, the plant life is, for the most part, in the form of one-celled green plants, called "algae," which, despite their microscopic size as individuals, far outweigh all forms of multicellular life put together.

(There are plants, notably mushrooms, that are not green. These are the "fungi." They cannot grow except on organic material and this makes them the equivalent of animals from the energy standpoint. That is, they, too, obtain their energy from a living or once-living source and not from the inanimate environment.)

It is the great attribute of the green plant that it does not actively eat anything in the sense that animals, or even fungi, do. The green plant can be grown in the absence of all organic material. They do, however, compete among themselves for light as animals do for food. Some plants can dispense with direct sunlight and grow in the shade, but no green plant can grow in complete darkness any more than any animal can long survive without food.

Anyone who has seen the plants of the tropical rain forests growing up toward the sun, casting out broad leaves to catch all the light they can, has seen a struggle for existence

as brutal and ruthless as any in which the animal kingdom engages; not a whit the less so because it is carried out in silence and slow motion, or because the insubstantial sunlight rather than living organisms are the target of the battle. Nor need one travel to the tropics to see this. The suburban lawn itself is witness to the same struggle, as the wide-spreading leaves of the dandelion plant catch the sun and remorselessly shade the slender blades of grass to death.

But sunlight is immaterial and, even if it supplies energy, the plant is made up of matter and must store its energy in the form of chemical substances. Where do these come from?

Since plants grow in the soil, there is the possibility that material composing the plant is drawn from that soil. In the 1660s this possibility was put to the test by a Flemish chemist named Jan Baptista van Helmont, who grew a willow tree in a pot for five years. The willow tree grew sturdily, gaining over 150 pounds in weight, yet the soil in which it grew lost only an ounce or two. Van Helmont's conclusion was that the plant was nourished not from the soil but from the water which he constantly added and without which any plant, however excellent the soil in which it grew, must die.

From our modern point of view, Van Helmont was right as far as he went. The soil consists largely of aluminum silicates and other compounds that play no role to speak of in metabolism. The function of soil is, for the most part, purely mechanical. It forms a solid, yet permeable, foundation in which the plant can root itself and in which water for plant use can be conserved.

The water, on the other hand, *is* necessary not merely because, as water, it makes up a major portion of plant tissue (as does the tissue of all forms of life), but because it dissolves small quantities of inorganic substances which the plant requires and which it cannot absorb except in solution. The use of fertilizers has as its purpose the restoration of these essential inorganics to soil depleted by previous generations of plants.

It is possible, in fact, to grow plants in a solution of essential inorganics with soil itself completely eliminated. This technique is known as "hydroponics." This certainly shows that Van Helmont was right in concluding that water was

more important than soil. Nevertheless, he was wrong to consider it *all*-important, for water and inorganics alone are insufficient to support life, as was made plain to the eighteenth-century chemists who began the careful study of organic substances.

The organics characteristic of living tissue contain carbon, and the carbon present in soluble form in fertile soil is insufficient to account for the increasing carbon of growing plant tissue. In hydroponics, in fact, the nutrient mixture can be completely free of carbon and the plant will nevertheless accumulate that element readily. If soil and water are eliminated as sources of carbon, then surely only the atmosphere is left as a possibility.

Now Van Helmont himself had been the first to discover the existence of different gases and, in fact, had described the properties of the one we now call carbon dioxide. Through the next century, research on the various gases continued at an intensifying pace and in 1727 the English physiologist Stephen Hales, who was interested in gases, found that the true solution to Van Helmont's problem of plant growth lay in the very carbon dioxide that Van Helmont had discovered!

But carbon dioxide makes up only 0.03 per cent of the volume of the atmosphere and it would be only natural to wonder if this seemingly minute amount can serve as carbon source for the myriad organisms all about us. However, the atmosphere is vast in volume and its total mass is unbelievably high to those of us who are used to clichés such as "light as air" and "an airy nothing." The total amount of carbon in the carbon dioxide of earth's atmosphere is, believe it or not, 4×10^{14} kilograms. Fifty times as much as this is dissolved in the ocean, either as carbon dioxide or as bicarbonate ion, so that the total mass of carbon available to life, in air and sea together, is 2×10^{16} kilograms.

It is estimated that the total amount of carbon in all living things on earth is 2.5×10^{14} kilograms; so that the carbon supply of sea and air is about eighty times the carbon existing in living tissue. As I said at the beginning of the chapter, the food must be at least ten times the mass of the eater, and if the carbon of carbon dioxide is considered the general food supply of all life, there is enough of it to spare to support all the individual organisms our planet now carries.

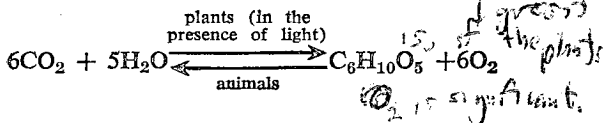
But it isn't all the same way!

In the 1770s the English Unitarian minister Joseph Priestley carried this one step further. He grew a plant under a bell jar and, as it grew, it incorporated the limited supply of carbon dioxide, available in the air about it, into its tissues and then could grow no more, despite the presence of sunlight and water. On the other hand, a mouse placed under a bell jar consumed the limited supply of oxygen and died. A mouse and a plant placed together under a bell jar, however, lived together in comfort far longer than either could by itself. From this, Priestley concluded that plants not only consumed carbon dioxide (to use modern terminology, which he, of course, did not) but gave off oxygen, while animals consumed oxygen and gave off carbon dioxide, each supplying what the other needed.

It was the Dutch physician Jan Ingenhousz who saw the significance of this on a planetary scale. In a book by him, published in 1779, he pointed out that the plant and animal life on earth formed a balance. The plants absorbed water and carbon dioxide and through the agency of light (something he was the first to point out clearly) formed its own tissues and gave off oxygen. Since light is needed for this formation of tissue out of carbon dioxide and water, the process is called "photosynthesis," which is "put together by light" in Greek. *How did plants live w/ animals?*

On the other hand, Ingenhousz went on, animals, eating plants and breathing oxygen, recombine plant tissue and oxygen and re-form carbon dioxide and water. *Surely their waste is*

With the development of food chemistry, it was recognized that starch was the usual form in which energy was stored by the plant. We can represent the empirical formula of starch as $C_6H_{10}O_5$. Ingenhousz's views can therefore be expressed by the equation:



Thus nothing material is used up. Carbon, hydrogen, and oxygen shuttle between plants and animals, and from land to sea in a process that is most commonly called "the carbon cycle." Other elements, too, are engaged in cyclic processes.

Nitrogen, sulfur, phosphorus, and so on are absorbed from the soil by the plants and incorporated into their tissue. Animals eat plants and make use of the various elements, then finally restore them to the soil in their droppings and in the form of their own bodies when death, followed by bacterial decomposition, arrives.

Only one thing is permanently used up in these cyclical chemical changes, and that is the energy of solar radiation. Far more of it is consumed than is sufficient to drive the cycle through its "uphill" stage of decreasing entropy, so that the net change is an increase in entropy, as the second law requires.

With the growing appreciation of catalysis that came about in the early nineteenth century, it was easy to see that some catalyst must be involved in photosynthesis. Suspicion rested almost at once on whatever it was that made plants green, for greenness was the one characteristic that all photosynthesizing plants had in common, and that which all non-photosynthesizing creatures lacked. (Some tropical birds have green feathers, but this is a different thing entirely.)

In 1819 two French chemists, Pierre Joseph Pelletier and Joseph Bienaimé Caventou, extracted the green substance of plants. They named it "chlorophyll" from Greek words meaning "green leaf," as straightforward a piece of naming as one can well imagine.

Not for an additional century, however, was the structure of chlorophyll worked out. Beginning in 1910, in a series of investigations carried out over a generation, the German chemists Richard Willstätter and Hans Fischer took the molecule of chlorophyll apart and deduced its composition from the nature of the fragments.

As it turned out, chlorophyll closely resembles, in its basic structural pattern, the molecule of heme, which is found in hemoglobin, catalase, and the cytochromes. Its chief points of difference are, first, that it contains an atom of magnesium in the center of the molecule, where heme contains an atom of iron. Secondly, attached to it is a long hydrocarbon molecule of a type known as "carotenoid."

The chlorophyll molecule is colored because it absorbs some of the wave lengths of visible light, but not others. In particular, it absorbs the long wave lengths of red and

orange regions of the spectrum, reflecting much of the rest. Sunlight, minus the red and orange, would be green in color, and it is that green that we see.

Chlorophyll, on absorbing light, is converted to a more energetic state. (In Chapter 11, I described the manner in which chlorine molecules are converted to the more energetic chlorine atoms by absorption of light. The process is by no means as simple in the case of chlorophyll and, indeed, is still not understood, but the principle remains—chlorophyll is raised to a high-energy state through the absorption of light.)

The light-activated chlorophyll can then expend its energy and return to "ground state" by bringing about some key, energy-consuming chemical reaction that is central to the process of photosynthesis.

But what is this "key chemical reaction"? The answer had to wait for the days of isotope tagging.

As you see from the equation on p. 356, photosynthesis involves the combination of carbon dioxide and water. Both these molecules contain oxygen and it might seem that the molecular oxygen produced by the green plant comes in part from both. After all, photosynthesis seems to be the reverse of respiration. In respiration, oxygen combines with organic substances to produce both carbon dioxide and water; therefore, in photosynthesis, should not both carbon dioxide and water break down to some extent to yield oxygen?

But reasoning, in the absence of experimental evidence, can be very misleading. To produce experimental evidence, a heavy but non-radioactive isotope of oxygen, O^{18} , was used by the American biochemists Samuel Ruben and Martin D. Kamen. (The most common oxygen isotope, O^{16} , makes up 99.76 per cent of all oxygen atoms, so that a compound consisting largely of O^{18} wherever oxygen atoms occur, stands out like a sore thumb in the mass spectrograph.)

In 1938, Ruben and Kamen prepared water containing O^{18} and grew plants in it. The oxygen produced had the same proportion of O^{18} as had the water. If, on the other hand, carbon dioxide tabbed with O^{18} were used, the oxygen produced contained only the usual very small amount of the heavy isotope. The conclusion was clear. It was the water molecule and the water molecule only that was split apart in photosynthesis and the oxygen produced came entirely

from the disrupted water molecule. The carbon dioxide molecule remained intact and was incorporated into the organic compounds built up by photosynthesis.

The main outline was thus made obvious. Sunlight was absorbed by chlorophyll and the energy so trapped was expended on breaking down water to hydrogen and oxygen, an energy-consuming reaction (see Fig. 75). This is the "photolysis" ("break down by light") of water.

The hydrogen so produced has two possible fates. Half the hydrogen atoms enter the respiratory chain, as can the hydrogen produced in an ordinary dehydrogenation. (In fact, the photolysis of water can be looked upon as the chlorophyll-catalyzed dehydrogenation of water.) Once that happens, hydrogen is reunited with oxygen at the final cytochrome oxidase step and water is formed again. However, in the process, three molecules of ATP have also been formed. Thus, the radiant energy of the sun has been converted into the chemical energy of ATP. (Since oxygen, produced along with the hydrogen that enters the respiratory chain, is consumed once more in forming water, there is no net production of oxygen here.)

The remaining half of the hydrogen atoms combine with carbon dioxide to form carbohydrates (and the oxygen formed along with it in this case is not needed and is discharged into the atmosphere). This is an energy-consuming reaction but takes place at the expense of the ATP formed by the photolysis and re-formation of water.

Thus the radiant energy of the sun, after being converted into the chemical energy of ATP, is used as the means for producing the carbohydrates (and, eventually, the lipids and all other tissue components) upon the energy content of which both plant and animal life maintain their existence.

Now it remains to determine the manner in which hydrogen combines with carbon dioxide. This proved to be a difficult nut to crack. When carbon dioxide carrying an unusual isotope is used, the isotope is found spread throughout the tissue components after only a short interval of time. Further, only intact "chloroplasts" (the plant cell particles that contain chlorophyll) can be used. Isolated chlorophyll showed no photosynthetic properties at all, so it was impossible to set up a simplified system for study as one could do for the various enzyme-catalyzed reactions.

The one hope was to use some radioactive isotope. Unfortunately, in the 1930s the only radioactive carbon isotope

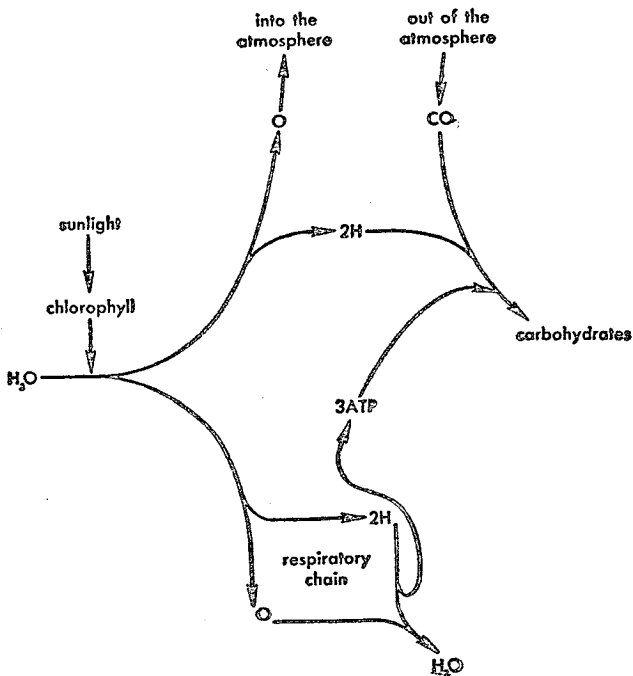


Figure 75. *Photosynthesis*

known was C^{11} , which was so unstable that it was mostly gone in about half an hour and was therefore very difficult to work with, and very expensive, for new supplies were constantly needed.

But then, in 1940, Ruben and Kamen discovered C^{14} , a radioactive isotope of carbon which, to everyone's surprise, proved comparatively stable and therefore easy to handle. (In fact, C^{14} is now undoubtedly the most useful of all the isotopes.)

After World War II the American biochemists A. A. Benson and Melvin Calvin applied C^{14} to photosynthesis. They

made use of a one-celled algae suspension and subjected it to carbon dioxide tabbed with C^{14} . After a short exposure they killed the cells, hoping that C^{14} would have a chance to take part in only a few initial reactions. The components of the killed cells were then separated by the technique of paper chromatography (see Chapter 16). It remained only to find out which spots on the paper were radioactive and to determine the chemical nature of the substances present in those spots.

After only a minute and a half, no less than fifteen different radioactive products could be isolated, and so Benson and Calvin had to reduce the time interval still further. Even after an exposure to radioactive carbon dioxide of only five seconds, five different compounds could be detected, with, however, most of the radioactivity present in two of them. Both of these were varieties of "phosphoglyceric acid," a three-carbon compound.

As a result of numerous such experiments, Benson and Calvin were able to work out a logical scheme of events. The carbon dioxide, as it enters the plant cell, is combined with "ribulose-1,5-diphosphate," a five-carbon sugar with two phosphate groups attached. (This compound has been referred to as the "carbon dioxide trap.")

The ribulose-1,5-diphosphate with the carbon dioxide added is now a six-carbon compound and promptly splits into two three-carbon compounds, the phosphoglyceric acid detected by paper chromatography. This phosphoglyceric acid next adds on the hydrogen produced by the photolysis of water to become a three-carbon sugar. This is the step that requires energy and makes use of the ATP produced by the photolysis and re-formation of water. The three-carbon sugar is then converted to starch by a number of steps (now worked out in detail) that require no further input of energy.

Perhaps the most remarkable thing about photosynthesis is its use of red and orange light as the source of energy. Photochemical reactions brought about in the laboratory usually require the far more energetic radiations at the blue end of the spectrum and into the ultraviolet. The ability of chlorophyll to make use of long-wave radiation is important, for sunlight is far richer in them than in ultraviolet. Furthermore, the long-wave radiation penetrates the dust

and haze of the atmosphere more efficiently than does the short-wave radiation, so that the former is a more reliable source of energy.

Photosynthesis may be unusually efficient for a photochemical reaction, too. In 1950, Otto Warburg and Dean Burk suggested that only four quanta of red light were necessary to supply the energy consumed by the fixation of one molecule of carbon dioxide in photosynthesis. Now the conversion of one mole of carbon dioxide to carbohydrate consumes 115 kilocalories, and the amount of red light necessary to effect such a conversion, if Warburg and Burk are correct, contains only 175 kilocalories. The efficiency is thus 115/175 or about 65 per cent, which is phenomenal.

The drama of life, then, plays itself out against the cascades of energy from the sun, as Mayer had suspected a century ago (see Chapter 4). The solar energy originates from the conversion of hydrogen to helium amid the incomprehensible heat and pressure of the sun's interior. The process is accompanied by a loss of mass and a corresponding appearance of energy.

The hydrogen atom loses less than 1 per cent of its mass in its conversion to helium, but over the huge volume of the sun, so many atoms are undergoing that conversion that the loss of matter amounts to 4,200,000 tons per second. The energy released as a result is equal to about 8×10^{21} (eight sextillion) kilocalories per second.

This release of energy through disappearance of mass is a spontaneous reaction *under the conditions of the solar interior* and represents a vast increase of entropy. All other radiating stars are likewise "running downhill" in this respect, some at a slower rate than the sun, some at a much faster rate.

The dominant process in the universe as a whole, then, is this entropy increase involved in stellar radiation. How the universe found itself in a position to support this entropy increase over the many billions of years in which it has been in existence and what will be the end of it (I mentioned the "heat-death" theory at the end of Chapter 5) is a matter for astronomers to determine. Biologists and biochemists can accept the energy flood from our own sun as a constant verity, at least over the space of time earthly life

has been in existence (perhaps two billion years) and for some additional billions of years to come.

It is only the increasing entropy represented by this solar energy that is responsible for the (much smaller) decreasing entropy associated with life on earth.

To be sure, most of the sun's energy is lost in space and wanders endlessly across the interstellar reaches to unknown destinations. A tiny fragment, however, is intercepted by the earth. Tiny it may be in comparison to the sun's total output, but it is still enormous, for it amounts to 2×10^{13} (twenty trillion) kilocalories per second. ~~$2 \times 10^{13} / 2 \times 10^{13} = 1$~~ $2 \times 10^{13} / 2 \times 10^{13} = 1$

But even this is only the amount potentially available. Nearly half is reflected by the cloud cover, the polar ice-caps, and the oceans. Of what is left, some is scattered by the atmosphere or is expended, without immediate use, in warming ocean water or desert sand.

Only 3 per cent of the total solar energy reaching the vicinity of the earth actually falls on green plants. Of this, perhaps two thirds is absorbed and comes within reach of chlorophyll. And, of the energy reaching the chlorophyll, perhaps two thirds is converted into high-energy phosphate groups.

That still means that 3×10^{11} kilocalories of high-energy phosphate bond is manufactured on earth each second through photosynthesis.

Assuming that an eater wastes 90 per cent of the energy content of the food he eats, the 3×10^{11} kilocalories per second produced by the plant world can support an animal life that utilizes 3×10^{10} kilocalories per second. About one eighth of all animal life, utilizing 4×10^9 kilocalories per second, lives, it is estimated, on earth's land surface.

The earth's population is now about 2.3×10^9 and if we suppose the average daily food intake to be 2000 kilocalories, then mankind is being supported by the expenditure of 53,000,000 kilocalories per second. More than 1 per cent of the mass of all animal life on land would thus have to be human.

It follows that an increase in the human population to eighty times the present value is all that the land vegetation can support, and even then, only if all other competing animal life is wiped out and if all plant life is dedicated to those forms man finds edible. If mankind makes use of the

Distance to sun's center ~ 93 million miles
Radius of earth ~ 4000 miles
 $\pi (4000^2) / 4 \times (93 \times 10^6)^2 = 4.3 \times 10^{-10}$
 ~~$4 \times (93 \times 10^6)^2 = 4.3 \times 10^{-10}$~~

See inside back cover

photosynthesis carried on by ocean plants, an increase in numbers to perhaps 650 times the present is potentially possible, again only if all animal life in the ocean is wiped out and man feeds on algae directly.

The maximum number of men, then, that can possibly be supported on earth by the available energy cascading out of the sun is 1.5×10^{12} ; that is, one and a half trillion.

10-10 An earth supporting that population would have just about one hundred square yards of land surface for each person on earth. There would be nearly fifty people per acre, even if Greenland, the Antarctic, and the Sahara Desert were thrown in. This compares with a current population density, today, of one and a half people per acre in Rhode Island, the most densely populated American state.

This may sound unbearable. What should seem even more unbearable, however, is the fact that if the present rate of human increase (a doubling of population every sixty years) is maintained this maximum population will be reached in about 550 years; that is, by 2500 A.D.

If, on the other hand, we grant that we will be unwilling to wipe out completely such animal life as cattle, swine, poultry, and fish (the meat of which forms such a desirable portion of the diet) and unable to wipe out completely such pests as rodents and insects, and that we must resign ourselves to a non-human animal life-mass equal to ten times our own, then the situation grows at once better and worse.

It grows better since our maximum population can then be only 150,000,000,000 and the population density over earth's land area will be down to five people per acre. There will be more elbow room and a more variegated diet. However, it will be worse, in that at the present rate of population growth mankind will reach this number in only 350 years, or by 2300 A.D.

To be sure, such computations don't take into account the fact that before such a state of affairs is reached the situation will be changed by the adoption of birth-control measures on a world-wide scale; or, perhaps, by the intervention of a disastrous nuclear war. Of the two, I, of course—as would any reasonable man—prefer the former, but in moments of pessimism I sometimes suspect the latter to be the more probable.

Unless man adapts himself to eat photosynthetic directly.

Yet it is wrong to despair. If man's future can be presented in dark colors, it can also be presented in bright ones. To some cosmic observer, watching the vast increase of entropy represented by the nuclear processes that feed the sun's radiation, the small jiggle of decreasing entropy introduced by life on earth (like a drop of spray shooting upward while Niagara plummets downward) would be completely unnoticeable.

And yet sheer quantity is not all. The complexity and versatility of life enforces a respect that cannot be elicited by raw sun power alone. And the human mind, which is the most elaborately intricate manifestation of life on earth, has a potentiality that has never yet been placed fully to the test.

If man is, and always will be, a slave to the laws of thermodynamics, he need not be an entirely helpless one. If he cannot subvert those laws, he can at least guide himself in such a way as to exact the utmost they will yield.

With the advent of nuclear power, man has taken a step that is comparable to the discovery of fire (with the discussion of which I began this book). Fire made man no longer directly dependent on the sun for energy and the atom may make man no longer even indirectly dependent on the sun.

To be sure, even with hydrogen fusion reactors complete and practical, he will still be dependent on the sun for biological energy; that is, for the production of foodstuffs through photosynthesis. But if man can learn enough about photosynthesis, the time may yet come when, with the energy of the fusion reactor, with the substance of the universe about him as raw material, and with a man-made catalyst equivalent to chlorophyll, he can create his own food independently of sun and green plant alike.

Free of any particular sun and planet, all space could be his home.

Or at least those portions of it that may not be already occupied by others.

Then we can fill up the earth to an arbitrary population density...

AUTHOR INDEX

- Amontons, Guillaume, 44
 Angstrom, Anders Jonas, 142
 Archimedes, 7, 8
 Arrhenius, Svante August, 137
 Aston, Francis William, 335
 Avogadro, Amadeo, 89

 Banting, Frederick Grant, 267
 Bartholin, Erasmus, 246
 Baumann, Eugen, 185
 Beaumont, William, 258
 Benda, C., 330
 Benson, A. A., 360
 Bernard, Claude, 264
 Bernoulli, Daniel, 62
 Berthelot, Pierre Eugène Marcelin, 92,
 106, 111
 Berthollet, Claude Louis, 111
 Berzelius, Jöns Jakob, 157, 164, 171,
 210, 222
 Best, Charles Herbert, 267
 Biot, Jean Baptiste, 248
 Black, Joseph, 42, 52
 Boerhaave, Hermann, 62
 Boltzmann, Ludwig, 66
 Boyle, Robert, 44
 Braconnot, M. H., 169, 209
 Brown, Robert, 71
 Buchner, Eduard, 196
 Burk, Dean, 362

 Cagniard de la Tour, Charles, 193
 Calvin, Melvin, 360
 Cannizzaro, Stanislao, 65
 Carnot, Nicolas Léonard Sadi, 53
 Cavendish, Henry, 167
 Caventou, Joseph Bienaimé, 357
 Celsius, Anders, 40
 Chance, Britton, 254
 Charles, Jacques Alexandre César, 45
 Chevreul, Michel Eugène, 170
 Clausius, Rudolf Julius Emanuel, 56,
 57, 66
 Clément, Nicolas, 154
 Copernicus, Nicolas, 17
 Cori, Carl F., 290
 Cori, Gerty, 290
 Cruikshank, William, 141

 Dakin, Henry Drysdale, 344
 Dalton, John, 64
 Daniell, John Frederic, 130
 Davy, Humphry, 64, 157
 Désormes, Charles Bernard, 154
 Dickens, F., 332
 Döbereiner, Johann Wolfgang, 157
 Dubrunfaut, Auguste Pierre, 196

 Eijkman, Christian, 256, 316
 Einstein, Albert, 51, 72, 148
 Embden, Gustav, 339

 Fahrenheit, Gabriel Daniel, 40
 Faraday, Michael, 25, 159
 Fischer, Emil, 211, 252
 Fischer, Hans, 357
 Frankland, Edward, 221
 Funk, Casimir, 256

 Galen, 166
 Galileo, 17, 19, 26, 28, 40, 44, 62
 Gay-Lussac, Joseph Louis, 169
 Gibbs, Josiah Willard, 115, 131
 Glauber, Johann Rudolf, 164
 Gomberg, Moses, 305
 Graham, Thomas, 203
 Gross, J., 187
 Guericke, Otto von, 20
 Guldberg, Cato Maximilian, 111
 Gutenberg, Johann, 17

 Hales, Stephen, 355
 Harden, Arthur, 255, 281, 320
 Helmholtz, Hermann Ludwig Ferdi-
 nand von, 48, 51, 115
 Henri V, 241
 Henry, Joseph, 25
 Henseleit, K., 345
 Herschel, William, 144
 Hertz, Heinrich Rudolf, 144
 Hess, Germain Henri, 101
 Hill, Archibald Vivian, 279
 Hofmann, Klaus, 218
 Hoppe-Seyler, Felix, 254
 Horstmann, August Friedrich, 115
 Huyghens, Christian, 142

 Ikhnaton, 13
 Ingenhousz, Jan, 356

Joule, James Prescott, 47, 65

Kamen, Martin D., 358
Katchalski, E., 212
Keilin, D., 325
Kekule, Friedrich August, 222
Kelvin, Lord, 45, 57, 65
Kendall, Edward Calvin, 187
Kirchhoff, Gottlieb Sigismund, 156, 169
Knoop, Franz, 333, 337
Kossel, Albrecht, 344
Krebs, Hans Adolf, 314, 345
Kühne, Wilhelm, 196

Langmuir, Irving, 159
Laplace, Pierre Simon de, 100
Lavoisier, Antoine Laurent, 50, 63, 100, 166, 173, 276, 299
Le Bel, Joseph Achille, 250
Le Châtelier, Henry Louis, 114, 115
Leeuwenhoek, Anton van, 192
Lewis, Gilbert Newton, 114, 116
Liebig, Justus von, 195, 222
Lind, James, 256
Lipmann, Fritz, 286, 317
Lohmann, K., 286
Lorentz, Hendrik Antoon, 144

Magnus-Levy, A., 187
Malus, Etienne Louis, 246
Marine, David, 185
Mariotte, Edme, 44
Martin, A. J. P., 215
Maxwell, James Clerk, 66, 143, 144, 145

Mayer, Julius Robert von, 48
Meckel, Johann Friedrich, 184
Mendeleev, Dmitri Ivanovich, 224
Menten, M. L., 241
Mering, J. von, 267
Meyerhof, Otto, 277
Michaelis, Leonor, 241, 304, 306
Minkowski, O., 267
Mirsky, A. E., 235
Mulder, Gerardus Johannes, 170

Nasse, O., 270
Nernst, Walther Hermann, 151
Newcomen, Thomas, 22
Newton, Isaac, 27, 28, 62, 141
Nicol, William, 247
Northrop, John Howard, 199

Ostwald, Wilhelm, 115, 156

Papin, Denis, 21
Pasteur, Louis, 193, 249, 276
Paul the Apostle, 192
Pauling, Linus, 235, 236, 306
Payen, Anselm, 169, 195
Pelletier, Pierre Joseph, 357
Persoz, J. F., 195
Pettenkofer, Max von, 174
Pfeffer, Wilhelm, 206
Phillips, Peregrine, 158
Pitt-Rivers, R. V., 187

Planck, Max Karl, 147
Priestley, Joseph, 356
Proust, Joseph Louis, 168
Prout, William, 171, 195

Quastel, J. H., 242

Rankine, William John Macquorn, 46
Réaumur, René Antoine Ferchault de, 46, 194
Regnault, Henri Victor, 174
Reiset, J., 174
Ritter, Johann Wilhelm, 144
Roemer, Olaus, 142
Roentgen, Wilhelm Konrad, 145
Rose, William C., 270, 271
Rouelle, G. F., 344
Ruben, Samuel, 358
Rubner, Max, 177, 178, 179, 182
Rumford, Count, 63

St. Martin, Alexis, 258
Sanger, Frederick, 216
Savery, Thomas, 22
Schmidt, C., 264
Schoenheimer, Rudolf, 336, 347
Schwann, Theodor, 181, 195
Smeaton, John, 23
Soddy, Frederick, 334
Sörensen, Sören Peter Lauritz, 206
Starling, Ernest Henry, 187
Sumner, James Batchellor, 199
Svedberg, Theodor, 208
Synge, R. L. M., 215
Szent-Györgyi, Albert, 313

Thompson, Benjamin, 63
Thomsen, Julius, 92
Thomson, William, 45, 65
Torricelli, Evangelista, 19
Tyndall, John, 205

Urey, Harold C., 336

Van der Waals, Johannes Diderik, 66
Van Helmont, Jan Baptista, 354, 355
Van't Hoff, Jacobus Hendricus, 115, 206, 250

Vesalius, Andreas, 17
Vigneaud, Vincent de, 219
Voit, Karl von, 174
Volta, Alessandro, 130

Waage, Peter, 111
Wallis, John, 36
Warburg, Otto, 312, 320, 323, 362
Ward, Joshua, 154
Watt, James, 23, 52
Wharton, Thomas, 185
Williamson, Alexander William, 113
Willstätter, Richard, 198, 357
Wirsung, Johann Georg, 183
Wöhler, Friedrich, 165, 195

Young, Thomas, 47
Young, W. J., 255, 281, 320

Zeeman, Pieter, 144

SUBJECT INDEX

- Absolute zero, 45
- Absorption, 261
- Acceleration, 27, 30
- Acetic acid, 153, 316
- Acetoacetic acid, 339, 347
- Acetone, 164
 - specific heat of, 42
- Acetylcoenzyme A, 318, 340
- Acid, 99
- Activation, energy of, 137
- Activity, 114
- Acyl phosphate, 288, 296
- Adenosine, 286
- Adenosine diphosphate, 288
- Adenosine monophosphate, 288
- Adenosine triphosphate, 286
- Adenylic acid, 288
- ADP, 288, 294
- Adrenocorticotrophic hormone, 218
- Aerobic dehydrogenase, 323
- Agriculture, 9
- Air pressure, 20
- Air pump, 20
- Alanine, 210, 238, 271, 347
- Albumen, 170
- Albuminous substances, 170
- Alcohol: ethyl, 278
 - combustion of, 157
 - heat of combustion of, 98, 176
 - specific heat of, 42
 - yeast and, 276
- Algae, 353
- Alimentary canal, 194, 258
- Alpha-ketoglutaric acid, 318, 332, 348
- Aluminum, 165
 - atomic number of, 124
 - specific heat of, 42
 - valence of, 221
- Amine group, 210
- Amino acids, 262
 - analysis of, 216ff.
 - arrangement of, 216, 217
 - combination of, 211, 213, 233
 - diet and, 270
 - discovery of, 210
 - essential, 272
 - non-essential, 272
 - optical activity of, 253
 - plasma albumin, 216
 - side chains of, 238
 - structural formula of, 233
 - tagged, 347
 - transamination of, 347
 - unnatural, 253
- Ammonia, 199, 209, 346
 - structural formula of, 222
- Ammonium cyanate, 165
- Ammonium ion, 126
- Ammonium nitrate, 107
- AMP, 288
- Amygdalin, 195
- Amylase, pancreatic, 260
- Amylase, salivary, 260
- Anabolism, 264, 275
- Anaerobic dehydrogenase, 323
- Anaerobic glycolysis, 276
- Angstrom unit, 142
- Angular momentum, 38
- Animals, domestic, 9
- Antibodies, 262
- Anti-ketogenic diet, 343
- Apoenzyme, 321
- Archimedes' principle, 7
- Arginase, 345
- Arginine, 210
 - breakdown of, 345
 - side chain of, 238
- Arsenic, 221
- Ascorbic acid, 253
- Aspartic acid, 210, 332, 346, 347
 - side chain of, 237
- Atmosphere, 355
- Atomic hydrogen, 81, 135
- Atomic nucleus, 123
- Atomic number, 123, 124
- Atomic oxygen, 81, 135
- Atomic theory, 64, 65
- Atomic weight, 83
- Atoms, 64
 - attraction among, 81
 - electrons in, 254
 - structure of, 123, 224
- ATP, 286
 - fatty acid catabolism and, 341-42
 - formation of, 297, 298, 329, 342
 - glycogen formation and, 293, 294
 - lactic acid catabolism and, 329
 - photosynthesis and, 358, 359
 - respiratory chain and, 328
 - urea cycle and, 345, 346

Avogadro's number, 89

Barometer, 20

Basal metabolic rate, 181

Base, 99

Benzene, 164, 306, 333

Benzoic acid, 337

Beriberi, 256

Beryllium, 165

Beta-lactoglobulin, 215

Beta oxidation, 340

Bile, 260

Bile salts, 260

Black body, 147

Blood glucose, 265

Blood iodine, 188

BMR, 181

thyroid abnormalities and, 186

Bomb calorimeter, 92

Bond, chemical, 81

coordinate, 284

covalent, 230

electrovalent, 229

hydrogen, 232

ionic, 127ff.

Boyle's law, 44

British thermal unit, 43

Bromine, 301

electron arrangement in, 226

valence of, 221

Bronze, 15

Brownian movement, 72

BTU, 43

Butyl alcohol, 215

Calcium, 300

atomic number of, 124

electronic symbol of, 227

valence of, 221

Calcium carbonate, 246

Calcium ion, 126

formation of, 229

Calcium oxide, 229

Caloric, 63

Calorie, 42, 48

Calorimetry, 92

Calx, 299

Camphor, 249

Cane sugar, 168

Carbohydrates, 171

digestion of, 260

RQ of, 174

Carbon, 15, 166, 299

asymmetric, 252

atomic number of, 124

atomic weight of, 83

combustion of, 85, 107

electron sharing and, 232

electronic symbol of, 227

heat of combustion of, 96

isotopes of, 335

life and, 355

organic substances and, 171

plants and, 346

radioactive isotopes of, 337, 360

symbol of, 84

tetrahedral atom of, 250, 251

valence of, 221

cycle, 356

Carbonate ion, 126

Carbon dioxide, 166, 173, 174, 199, 263, 277, 301, 335, 346

formation of, 85

heat of formation of, 96

molecular weight of, 85

photosynthesis and, 358, 359

plant life and, 355

structural formula of, 222

tissue production of, 312 ff., 316

yeast production of, 276

Carbon dioxide trap, 361

Carbon monoxide, 102

platinum and, 159

Carbonic anhydrase, 257

Carboxyl group, 210, 211, 288, 313

Carotenoid, 357

Catabolism, 261, 275

Catalase, 308, 325

turnover number of, 239

Catalysis, 150

life and, 191 ff.

thermodynamics and, 160, 161

Catalyst, 158

organic, 195

protein, 199

Cell, 329

enzymes and, 239

Cell nucleus, 329

Cellulose, 169, 212, 260

Cell, voltaic, 130

Celsius scale, 40

Centigrade scale, 40

Centimeter, 30

Centrifugal effect, 208

Chain reaction, 150

Charcoal, 16

Charge, electric, 123

Charles' law, 45

Chemical bonds, 81

Chemical cell, 130

Chemical element, 83

Chemical energy, 77

Chemical equation, 86

Chemical equilibrium, 107, 109

free energy and, 121

Chemical kinetics, 133

Chemical potential, 106, 117

Chemical reactions, 82

electrons and, 124

temperature and, 134 ff.

Chemical thermodynamics, 115

Chloride ion, 125

formation of, 228

Chlorine, 87, 300

atomic number of, 124

atomic weight of, 83

electron arrangement in, 125, 226

electronic symbol of, 227

electron sharing and, 229

electron transfer and, 228

hydrogen combination with, 141

light and, 149 ff.

molecular weight of, 88

symbol of, 83

valence of, 221

Chlorine molecule, 232

Chlorophyll, 357

- Chloroplast, 359
 Cholesterol, 343
 Chromatography, paper, 215
 Chromium, 124
 Chymotrypsin, 261
 Citric acid, 315, 318, 340
 Citric acid cycle, 315
 Citrulline, 345
 Coal, 176
 Cobalt, 257
 atomic number of, 124
 Coenzyme A, 317
 fatty acid catabolism and, 339
 Coenzymes, 255
 vitamins and, 256
 Colloids, 203
 particle size of, 205
 Combustion, 77, 298
 spontaneous, 133
 Compass, 17
 Competitive inhibition, 243, 244, 314
 Compound, 85
 Compound ion, 126
 Compound lipid, 343
 Concentration, 112
 Condensation, 274
 Conjugate proteins, 254
 Contact process, 158
 Coordinate bond, 284
 Copper, 15, 257, 299, 324
 atomic number of, 124
 electrons and, 127, 128
 specific heat of, 42
 symbol of, 127
 valence of, 221
 Copper sulfate, 127
 Cori ester, 290
 Corrosive sublimate, 244
 Coupled reaction, 280
 Covalent bond, 230
 Cozymase, 255, 320
 Cretin, 185
 Crystalloids, 203
 Cupric ion, 126, 324
 Cuprous ion, 126, 324
 Cyanide group, 326
 Cysteine, 210, 220, 347
 Cystine, 210, 270
 Cytochromes, 325
 oxidation potential of, 326
 poisoning of, 326
 Cytoplasm, 329

 Daniell cell, 130
 Decarboxylation, 278
 oxidative, 315
 Dehydration, 303
 Dehydrogenase, 321
 Dehydrogenation, 296, 300, 301, 303,
 316, 327
 catalysis of, 321
 fatty acid catabolism and, 339
 high-energy phosphate bonds and,
 317
 Krebs cycle and, 318, 319
 Denatured protein, 219
 hydrogen bonds and, 234, 235
 Deuterium, 336

 Dextrin, 260
 Diabetes mellitus, 266
 ketone bodies and, 339
 ketosis and, 343
 Dialysis, 204
 molecular weight and, 206
 Diastase, 195
 Dicarboxylic acids, 314
 Dichloromethane, 250
 Dickens shunt, 332
 Dietary deficiency disease, 256
 Digestion, 194, 258
 Diphosphoglyceric acid, 296
 Diphosphopyridine nucleotide, 320
 Disorder, 73 ff.
 Döbereiner's lamp, 157
 Double bond, 303
 Double refraction, 246
 DPN, 320
 oxidation potential of, 327
 Drunkard's walk, 72
 D-series, 252
 Ductless gland, 184
 Dynamic equilibrium, 109
 Dyne, 31
 Egg albumin, 106
 Egg white, 170
 Electrical cell, 130
 Electrical energy, 130
 Electric charge, 123
 Electricity, 24, 47
 Electric potential, 129
 Electromagnetic radiation, 143
 Electromagnetic spectrum, 145
 Electromotive force, 129
 Electron, 25, 123
 chemical reactions and, 124
 donation of, 283
 electric charge of, 123
 light and, 144
 mass of, 129
 oxidation-reduction and, 300
 resonance theory and, 306
 sharing of, 229
 shells of, 226
 unpaired, 303
 valence, 230
 Electronics, 130
 Electron microscope, 331
 Electrovalent bond, 229
 Elements: chemical, 82
 atomic weights of, 83
 electronic symbols of, 227
 periodic table of, 224, 225
 symbols of, 83
 Elephant, 182
 EMF, 129
 Emulsin, 195
 Endocrine gland, 184
 Endothermic reaction, 106
 Energy, 4
 activation, 138, 155, 156
 chemical, 77
 conservation of, 33 ff., 47, 48
 difference of, 34, 93
 electrical, 130
 free, 116

Ecology 35

328, 329, 342, 350, 362, 36

- heat and, 46 ff.
- inanimate, 10 ff.
- kinetic, 26, 32
- light, 145
- mass and, 51
- mechanical, 33, 34
- potential, 33
- rate of expenditure of, 5, 180
- solar, 48, 51, 353, 361
- sources of, 76 ff., 353, 354
- temperature differences and, 52
- tissue utilization of, 350
- Enthalpy, 94
- Entropy, 57 ff.
 - chemical reactions and, 116
 - decrease of, 71
 - disorder and, 73 ff.
 - increase of, 58
 - life and, 189 ff.
 - optical activity and, 250
 - protein denaturation and, 220
 - protein formation and, 214
 - sun and, 362
- Enzyme, 196
 - catalytic activity of, 237 ff.
 - crystalline, 199
 - dialysis of, 255
 - heat and, 198
 - heme and, 254
 - hydrogen bonds and, 238
 - inactivation of, 198
 - kinetics of, 240 ff.
 - life and, 200 ff.
 - metal atoms in, 324
 - molecular weight of, 208
 - nature of, 197 ff.
 - optical activity and, 254
 - properties of, 202 ff.
 - reaction rates and, 240
 - specificity of, 238
 - substrate and, 240
 - trace minerals and, 257
- Enzyme-substrate complex, 240
- Equations, chemical, 86
- Equilibrium: chemical, 107
 - free energy and, 121
- Equilibrium constant, 113
 - free energy and, 121, 122
- Erg, 31, 48
- Essential amino acids, 272
- Ethylene, 103
- Exothermic reactions, 106
- Explosions, 138 ff.
- FAD, 323
 - oxidation potential of, 327
- Fahrenheit scale, 40
 - absolute zero and, 46
- Fats, 170
- Fatty acids, 260, 262
 - anabolism of, 341
 - catabolism of, 339, 340
 - phenyl derivatives of, 334
 - tagged, 336
- Feedback, 266
- Ferment, 195
- Fermentation, 192
- Ferric chloride, 222
- Ferric ion, 126, 324
- Ferrous chloride, 222
- Ferrous ion, 126, 324
- Fibrinogen, 208
- Fibroin, 218
- Fire, 11
 - air and, 166
 - man and, 14 ff.
 - metallurgy and, 15
 - muscular effort and, 18 ff.
 - worship of, 13
- First law of thermodynamics, 53
- Fistula, 258
- Flavin, 323
- Flavin adenine dinucleotide, 323
- Flavin mononucleotide, 323
- Flavo-enzymes, 323
- Fluoride ion, 126
- Fluorine, 300
 - atomic number of, 124
 - electron arrangement in, 227
 - valence of, 221
- FMN, 323
- Food, 167, 170
- Food chain, 351
- Force, 7, 27
 - units of, 31
- Free energy, 116
 - chemical equilibrium and, 121
 - electric cells and, 131
 - equilibrium constant and, 121, 122
 - hydrogen combustion and, 135 ff.
 - molar, 118
- Free radical, 303
 - stabilization of, 305 ff.
- Frequency, light, 142
- Friction, 38
- Fructose, 169, 196, 212, 260
- Fructose diphosphate, 281, 296
- Fruit sugar, 168
- Fuel cell, 131
- Fumaric acid, 313
- Fungi, 353
- Fusion, heat of, 43
- Galactose, 244, 261
- Gamma rays, 145
- Gases, 40
 - expansion of, 65
 - kinetic theory of, 66
 - pressure and, 44
 - temperature and, 40, 44
- Gastric juice, 195, 260
- Gelatin, 209, 270
- Generator, electric, 25
- Gibbs function, 116
- Gland, 183, 184
- Glauber's salt, 164
- Glucagon, 266
- Glucose, 156, 169, 196, 212, 244, 260, 264
 - blood content of, 264
 - combustion of, 173
 - energy of glycolysis of, 277, 278
 - fermentation of, 276
 - formation of, 280, 281
 - formula of, 169
 - heat of combustion of, 175, 176

- lactic acid from, 277
- urine content of, 266
- Glucose-1-phosphate, 291
- Glucose-6-phosphate, 292
- Glutamic acid, 210, 332, 346, 347
 - side chain of, 238
- Glyceraldehyde, 252
- Glyceraldehyde-3-phosphate, 296
- Glycerol, 260, 342
- Glycine, 210, 337
- Glycogen, 264, 277
 - ATP and, 293, 294
 - body supply of, 267
 - phosphorolysis of, 291
- Glycogenesis, 294
- Glycogenolysis, 294
- Glycolysis, 283
 - ATP formation in, 297, 298
- Goiter, 185
- Gold, 14
 - atomic number of, 124
 - valence of, 221
- Gram, 32
- Gram atomic weight, 90
- Gram molecular weight, 89
- Grape sugar, 168
- Gravitation, universal, 29
- Greek fire, 16
- Green plants, 358
- Gunpowder, 16
- Harden-Young ester, 281, 296
- Heat, 38 ff.
 - chemical reactions and, 91 ff.
 - disorder and, 73
 - energy and, 47
 - flow of, 39, 56 ff., 70
 - as fluid, 62
 - latent, 43
 - light and, 146
 - mechanical equivalent of, 48
 - as motion, 62 ff.
 - nature of, 61 ff.
 - reaction, 91
 - specific, 42
 - units of, 42, 48
 - work and, 47
- Heat content, 39, 94
- Heat-death, 60
- Heavy hydrogen, 336
- Heavy nitrogen, 336
- Heavy water, 336
- Helium, 52, 362
 - atomic number of, 124
- Heme, 254, 326
- Heme enzymes, 254, 326
- Hemocyanin, 208
- Hemoglobin, 213, 254, 326
 - molecular weight of, 206
- Hepatic vein, 265
- Hess's law, 102
- High-energy phosphate bond, 286
 - dehydrogenation and, 303
 - production of, 363
- Histidine, 210
 - side chain of, 238
- Homogenates, 289
- Hormones, pancreatic, 266
 - pituitary, 217
 - thyroid, 187
- Hyaluronic acid, 212
- Hydration, 303
- Hydrochloric acid, 195, 261
- Hydrogen, 51, 81, 87, 296
 - acceptor molecules for, 307, 320, 322
 - atomic number of, 124
 - atomic weight of, 83
 - catalyzed combustion of, 160
 - chlorine combination with, 141
 - combustion of, 85, 106, 118, 131, 135 ff., 157, 166
 - electronic symbol of, 227
 - electron-sharing of, 230
 - free energy of combustion of, 131
 - heat of combustion of, 97, 176
 - iodine and, 108, 119, 120
 - isotopes of, 336
 - oxidation potentials and, 310
 - photosynthesis and, 359, 360
 - radioactive isotope of, 337
 - reducing ability of, 299
 - solar consumption of, 362
 - symbol of, 83
 - valence of, 221
- Hydrogenation, 300
- Hydrogen bond, 232
 - enzymes and, 238
 - proteins and, 235
- Hydrogen carriers, 320
 - chain of, 326
- Hydrogen chloride, 87
 - heat of formation of, 99
 - structural formula of, 222
- Hydrogen iodide, 108, 119, 120
- Hydrogen ion, 126, 301
- Hydrogen peroxide, 239, 308, 325
- Hydrolysis, 261, 262
- Hydroponics, 354
- Hyperthyroidism, 187
- Hypothyroidism, 187
- Ice, 43, 78
- Iceland spar, 246
- Industrial Revolution, 24
- Infrared radiation, 1¹
- Inhibition, competitive, 24⁷, 243, 314
- Inorganic substances, 164
- Insulin, 266
 - amino acid arrangement in, 217
 - isolation of, 267
- Intermediary metabolism, 262
- Intestinal juice, 260
- Inulin, 212
- Invertase, 195
- Iodide ion, 126
- Iodine, 108, 119, 120
 - atomic number of, 124
 - atomic weight of, 83
 - BMR and, 187
 - electron arrangement in, 226
 - protein-bound, 188
 - symbol of, 83
 - thyroid and, 185
 - valence of, 221

- Iodized salt, 186
- Ionic bond, 126 ff.
- Ionization, 301
- Ions, 125
 - mass spectrograph and, 335
- Iron, 15, 299, 308
 - atomic number of, 124
 - atomic weight of, 83
 - enzymes and, 324
 - hemoglobin and, 254
 - symbol of, 83
 - valence of, 221
- Iron ore, 299
- Isoleucine, 210, 238, 244
 - dietary requirement of, 272
- Isomer, 222
 - number of, 223
- Isoprene, 212
- Isotopes, 334
 - radioactive, 337
- Joule, 32, 48
- Joule-Thomson effect, 65
- Kelvin scale, 46
- Keto acid, 315
- Ketogenic diet, 343
- Keto group, 315, 339
- Ketone bodies, 339
- Ketosis, 343
- Kilocalorie, 43
- Kilogram, 30
- Kinetic energy, 26, 32
 - molecules, 69 ff.
- Kinetic theory of gases, 66
- Kinetics, chemical, 133
- Krebs cycle, 315
 - amino acid catabolism and, 347
 - dehydrogenations in, 318, 320
 - fatty acid catabolism and, 341
 - intermediates of, 332
 - mitochondria and, 330
- Lactase, 260
- Lactic acid, 277
 - breakdown of, 280 ff., 315, 330
- Lactic acid dehydrogenase, 321
- Lactose, 169, 260
- Latent heat, 43, 79
- Lavoisier-Laplace generalization, 100, 101
- Lawrencium, 124
- Lead, 42
 - atomic number of, 124
- Lead-chamber process, 154
- Leaven, 192
- Le Châtelier's principle, 114
- Leucine, 210, 238, 244, 347
 - dietary requirement of, 272
- Lever, 7
- Life: catalysts and, 191 ff.
 - definition of, 200
 - distinction from non-life of, 2 ff., 162 ff., 189 ff., 200 ff.
 - entropy and, 60, 189 ff.
 - enzymes and, 200 ff.
 - fermentation and, 193
 - thermodynamics and, 151 ff., 177 ff.
- Light, 142
 - chemical reactions and, 141
 - chlorine and, 149 ff.
 - chlorophyll and, 357
 - double refraction of, 246
 - electrons and, 144
 - energy of, 145
 - frequency of, 142
 - heat and, 146
 - photosynthesis and, 361
 - plants and, 353
 - polarized, 246
 - refraction of, 246
 - scattering of, 205
 - velocity of, 142
 - wavelengths of, 142
- Lipase, 260
- Lipid metabolism, 342, 343
- Lipids, 170
 - body stores of, 268
 - catabolism of, 333 ff.
 - combustion of, 173
 - digestion of, 260
 - formula of, 170
 - heat of combustion of, 175, 176, 177
 - RQ of, 174
- Liver, 184
 - glucose and, 265, 266
 - glycogen in, 265, 267
 - secretion of, 260
- Low-energy phosphate bond, 286
- L-series, 252
- Lymph gland, 184
- Lysine, 210
 - dietary requirement of, 272
 - side chain of, 238
- Machine, 7
- Magdeburg hemispheres, 21
- Magnesium, 300, 357
 - atomic number of, 124
 - atomic weight of, 83
 - electronic symbol of, 227
 - symbol of, 83
 - valence of, 221
- Magnesium ion, 126
- Magnetism, 17, 47
- Malic acid, 313
- Malonic acid, 242, 313
- Maltase, 260
- Maltose, 195, 260
- Man, 3 ff.
 - air and, 166
 - animals and, 9
 - BMR of, 181
 - energy and, 25
 - fire and, 11 ff., 18 ff.
 - food supply of, 352
 - inanimate energy and, 11 ff.
 - longevity of, 183
 - maximum population of, 363
 - RQ of, 174, 177 ff.
 - tools and, 6
 - work and, 3
- Manganese, 257
 - atomic number of, 124
- Manometer, Warburg, 312

- Mariotte's law, 44
- Mass, 28
 - conservation of, 51
 - energy and, 51
 - weight and, 30
- Mass action, law of, 111
- Mass-energy, conservation of, 52
- Mass spectrograph, 335
- Maxwell-Boltzmann distribution law, 67
- Maxwell's demon, 71
- Mechanical energy, 33
- Membranes, semi-permeable, 203
- Mercury, 119
 - atomic number of, 124
 - expansion of, 44
 - thermometer and, 40
 - valence of, 221
- Mercury bichloride, 244
- Metabolic disease, 266
- Metabolic intermediate, 281
- Metabolism, 181, 258
 - disorder of, 266
 - intermediary, 262
- Metal, 14
 - specific heat of, 42
- Metallurgy, 15
- Meter, 30
- Methane, 230
 - boiling point of, 232
 - free energy of combustion of, 131
 - heat of combustion of, 98, 176
 - melting point of, 232
 - structural formula of, 222
- Methionine, 210, 270
 - dietary requirement of, 272
- Methylene blue, 322
- Metric system, 30
- Michaelis-Menten equation, 242
- Micron, 142
- Microwaves, 145
- Milk, 277
- Milk sugar, 166
- Millimicron, 142
- Minerals, trace, 257
- Mitochondria, 330, 343
- Molar free energy, 118
- Molar heat of formation, 105
- Molar heat of reaction, 96
- Molar heat of solution, 99
- Mole, 89
- Molecular weight, 85
 - osmotic pressure and, 206, 207
 - sedimentation rate and, 207
- Molecules, 65
 - attraction among, 77 ff.
 - giant, 208
 - non-polar, 231
 - polar, 231
 - velocity of, 67 ff.
- Molybdenum, 257, 324
- Momentum, conservation of, 37
- Moon, 28
- Motion, 27
 - energy and, 26 ff., 32
 - heat and, 63 ff.
 - perpetual, 52
- Motor, electric, 25
- Mouse, 182
- Muscle glycogen, 267
- Muscle glycolysis, 277
- Muscle lactic acid, 277, 278, 279
- Mushrooms, 353
- Native protein, 219
- Negative ion, 125
- Negative nitrogen balance, 271
- Neutralization, heat of, 99
- Neutrons, 123
- Newton, 31
- Nickel, 165
 - atomic number of, 124
- Nicol prism, 247
- Nicotinamide, 321
- Nicotinic acid, 321
- Nitrate ion, 126
- Nitrogen, 166, 167, 178, 347
 - atomic number of, 124
 - atomic weight of, 83
 - electronic symbol of, 227
 - heavy, 336
 - isotopes of, 336
 - symbol of, 83
 - valence of, 221
- Nitrogen balance, 270
- Nitrogen dioxide, 155
- Non-essential amino acids, 272
- Non-life and life, 1 ff., 162 ff., 189 ff., 199 ff.
- Non-polar molecules, 231
- Nucleic acids, 171, 286
- Nucleotide, 320
- Nucleus, atomic, 123
- Nucleus, cell, 329
- Obligate anaerobes, 326
- Oil, 170
- Optical activity, 249
- Organelles, 329
- Organic acid, 288
- Organic substances, 164
 - synthesis of, 165, 172
- Organized ferments, 196
- Ornithine, 345
- Orthophosphoric acid, 284
- Osmosis, 206
- Osmotic pressure, 206
- Oxaloacetic acid, 313, 317, 332, 340, 342, 348
- Oxidase, 327
- Oxidation, 298
- Oxidation potential, 310
 - of hydrogen carriers, 327
- Oxidation-reduction reactions, 301
- Oxidative decarboxylation, 315
- Oxidative phosphorylation, 328
- Oxide, 299
- Oxide ion, 229
- Oxidizing agent, 299
- Oxygen, 15, 16, 81, 106, 118, 132, 135, 153, 157, 173, 174, 239, 298, 322
 - atomic number of, 124
 - atomic weight of, 83
 - electron sharing of, 231
 - electron symbol of, 227, 307
 - hydrogen acceptance by, 327

- isotopes of, 358
- lack of, 326
- life and, 166, 276
- molecule of, 307
- platinum and, 160
- production by plants of, 356
- symbol of, 83
- tissue absorption of, 311 ff.
- valence of, 221
- Oxygen debt, 279
- Oxytocin, 217
- Ozone, 84
- Palladium, 157
- Pancreas, 183
 - hormones of, 266
- Pancreatic amylase, 260
- Pancreatic juice, 260
- Pantothenic acid, 317
- Paper, 133
- Paper chromatography, 215, 361
- Para-aminobenzoic acid, 245
- PBI, 181
- Pepsin, 195, 261
 - crystallization of, 199
- Peptidase, 261
- Peptide, 211
 - optical activity and, 253
- Peptide link, 212
 - heat of hydrolysis of, 274
 - structural formula of, 234
- Periodic table, 224, 225
- Perpetual motion, 52
- Phenylacetic acid, 337
- Phenylalanine, 210, 238
 - dietary requirement of, 272
- Phlogiston, 62, 77
- Phosphate bonds, 286
- Phosphate ion, 126, 284
- Phosphine, 283
- Phosphoglyceric acid, 361
- Phosphoric acid, 284
- Phosphorolysis, 290
- Phosphorus, 167
 - atomic number of, 124
 - atomic weight of, 83
 - electronic symbol of, 227, 283
 - symbol of, 83
 - valence of, 221
- Photolysis, 356
- Photosynthesis, 356
 - efficiency of, 362
- Pituitary gland, 217
- Planck's constant, 148
- Plane-polarized light, 246
- Plasma albumin, 216
- Plastein, 274
- Platinum, 124
 - catalytic activity of, 158, 237
 - combustion and, 157
- Plutonium, 124
- Poisons, 244
- Polarimeter, 247
- Polarized light, 246
- Polar molecule, 231
- Polaroid, 247, 248
- Polonium, 50
- Population, 363
- Portal vein, 265
- Positive ions, 125
- Positive nitrogen balance, 271
- Potassium, 300
 - atomic number of, 124
 - atomic weight of, 83
 - electronic symbol of, 227
 - symbol of, 83
 - valence of, 221
- Potassium acid phosphate, 281
- Potassium chromate, 300
- Potassium cyanide, 326
- Potassium ion, 126
- Potassium nitrate, 16, 154
- Potassium permanganate, 300
- Potential energy, 33
- Power, 5
- Primary phosphate ion, 284
- Printing, 17
- Proline, 210
- Prontosil, 244, 245
- Proportionality constant, 112, 113
- Prosthetic group, 254
- Proteases, 261
- Protein, 171
 - amino acid arrangements in, 213 ff.
 - amino acids of, 210
 - breakdown of, 209
 - catabolism of, 344
 - catalytic, 199
 - conjugate, 254
 - cysteine and, 220
 - denaturation of, 220
 - dialysis of, 203
 - dietary essentiality of, 269
 - digestion of, 260
 - entropy and, 214
 - enzymes as, 198
 - formula of, 214
 - heat of combustion of, 177
 - helical structure of, 236
 - hydrogen bond and, 234
 - molecular weights of, 205 ff.
 - native, 219
 - RQ of, 174
 - sedimentation rate of, 207
 - strength of, 219
 - synthesis of, 217, 274
 - Tyndall effect and, 205
 - wear and tear of, 271
 - x-ray and, 235
- Protein-bound iodine, 188
- Protons, 123
 - electric charge of, 123
- Protozoa, 192
- Prussic acid, 326
- Pumping, 18
- Pyridine, 255
- Pyridino-enzyme, 321
- Pyridoxal phosphate, 347
- Pyridoxine, 347
- Pyrophosphate bond, 286
- Pyruvic acid, 315, 352
- Quanta, 148
- Quantum theory, 148, 149
- Quartz, 248

- Radioactivity, 50
Radio waves, 145
Radium, 50
 atomic number of, 124
Rankine scale, 46
Reactions, chemical, 82
 electrons and, 124
 endothermic, 106
 exothermic, 106
 heat of, 91 ff.
 spontaneous, 106 ff., 117
 temperature and, 134 ff.
Reagents, test, 220
Réaumur scale, 46
Redox reactions, 301
Reduction, 299
Relativity, theory of, 50
Resonance theory, 306
Respiratory chain, 326
 ATP production and, 329
Respiratory quotient, 173
Reversible reaction, 108
Riboflavin, 323
Ribonuclease, 217
Ribulose-1,5-diphosphate, 361
Rice, 256
RQ, 173, 178 ff.
Rubber, 212
- Saliva, 260
Salivary amylase, 260
Salt, 87, 126
 iodized, 186
Saltpeter, 16, 154
Scurvy, 256
Second, 30
Secondary phosphate ion, 284
Second law of thermodynamics, 55
Sedimentation rate, 205
 proteins and, 207
Serine, 210, 347
 side chain of, 238
Silicon, 221
Silk, 219
Silver, 144
 atomic number of, 124
 symbol of, 223
Silver fulminate, 223
Silver isocyanate, 222
Silver nitrate, 144
Simple lipids, 343
Sodium, 87, 300
 atomic number of, 124
 atomic weight of, 83
 electronic symbol of, 227
 electrons of, 125
 electron transfer and, 228
 radioactive isotope of, 337
 symbol of, 83
 valence of, 221
Sodium ammonium tartrate, 249
Sodium carbonate, 100
Sodium chloride, 126, 127, 228, 300
 formation of, 87
 heat of formation of, 99
Sodium hydroxide, 99
Sodium ion, 125
 formation of, 228
- Sodium sulfate, 164
Solar energy, 353, 362
Solution, heat of, 99
Specific heat, 42
Spectrum, electromagnetic, 146
Spectrum light, 141
Spontaneous combustion, 133
Stains, cell, 329
Starch, 168, 212, 356
 breakdown of, 156
 digestion of, 260
Steam, 21 ff.
Steam engine, 22 ff.
 efficiency of, 52
Stearic acid, 342
Stearin, 342
Stomach, 259
 juices of, 194, 195
Streptomycin, 253
Subatomic particles, 123
Substrate, 238
Substrate concentration, 241
Succinic acid, 242, 313, 323
Succinic acid dehydrogenase, 323, 326, 328
Sucrase, 260
Sucrose, 169, 196, 197, 260
Sugar, 168, 169
 optical activity of, 252, 253
 unnatural, 253
Sugar of gelatin, 209
Sugar phosphate, 285, 290
Sulfanilamide, 244, 245
Sulfate ion, 126
Sulfide ion, 126
Sulfur, 15, 16, 167, 300
 atomic number of, 124
 atomic weight of, 83
 combustion of, 153
 electronic symbol of, 227
 symbol of, 83
 valence of, 221
Sulfur dioxide, 153
 combustion of, 158
Sulfuric acid, 153, 156
 formation of, 152 ff.
Sulfurous acid, 153
Sulfur trioxide, 154, 158
Sun, 362
 energy of, 48, 51
 worship of, 13
Surface catalysis, 159
System, closed, 58
- Temperature, 39
 absolute, 45
 chemical reactions and, 134 ff.
 molecular energy and, 69 ff.
 scales of, 46
Tertiary phosphate ion, 284
Test reagents, 220
Thermochemistry, 102
Thermodynamics, 53
 catalysis and, 160, 161
 chemical reactions and, 100 ff., 115
 electrical energy and, 131
 ferments and, 195
 life and, 151 ff., 177 ff.

- Thermometer, 40
- Thiamine, 315
- Thiol group, 219, 318
- Thorium, 124
- Threonine, 210, 347
 - dietary requirement of, 272
 - discovery of, 270
 - side chain of, 238
- Thyroid gland, 185
 - BMR and, 187
 - hormones of, 187
- Thyroxine, 187
- Tin, 124
- Tissue extracts, 290
- Tools, 6
- Torriceillian vacuum, 20
- Tortoise, 183
- TPN, 321
- Trace minerals, 257
- Tracer, 335
- Transaminase, 347
- Tricarboxylic acid cycle, 315
- Tri-iodothyronine, 187
- Triphenyl methyl, 305
- Triphosphopyridine nucleotide, 321
- Triphosphoric acid, 286
- Trypsin, 261
- Tryptophan, 210, 270
 - dietary requirement of, 272
- Tungsten, 124
- Turnover number, 239
- Turpentine, 249
- Two-carbon fragment, 318
- Tyndall effect, 205
- Tyrosine, 211, 270
 - side chain of, 238
- Ultracentrifuge, 208, 330
- Ultramicroscope, 205
- Ultraviolet radiation, 144
- Units of measurement, 29 ff.
- Unnatural amino acids, 253
- Unnatural sugars, 253
- Unorganized ferments, 196
- Uranium, 50, 145
 - atomic number of, 124
- Urea, 263, 344
 - breakdown of, 199
 - formation of, 178
 - synthesis of, 165
- Urea cycle, 345
- Urease, 199
- Urine, 266
- Vacuum, 20
 - steam and, 21
- Valence, 221
 - periodic table and, 224, 225
- Valence electrons, 230
- Valine, 210, 238, 244, 347
 - dietary requirement of, 272
- Vaporization, heat of, 43
- Vapor pressure, 79
- Velocity, 30
- Virus, 208
- Vital force, 165, 191, 194
- Vitamin, 256
- Vitamin B₁, 315
- Volt, 131
- Voltaic cell, 130
- Warburg manometer, 312
- Warburg's old yellow enzyme, 323
- Washing soda, 99
- Water, 153, 239, 263, 308, 346
 - boiling of, 79
 - dissociation of, 81
 - expansion of, 44
 - formation of, 85, 166
 - formula of, 82
 - free energy of formation of, 131
 - heat of formation of, 97
 - heat of fusion of, 43
 - heat of vaporization of, 43
 - heavy, 336
 - molecular weight of, 82
 - photolysis of, 359
 - plants and, 354
 - polar properties of, 231, 232
 - specific heat of, 42
 - structural formula of, 222
 - tissue formation of, 316
 - vaporization of, 79
- Water power, 18
- Wave length, 142
- Weight, 30
- Whale, 183
- Wood, 167
- Work, 3
 - heat and, 47
 - units of, 31
- X rays, 145
 - protein structure and, 235
- Yeast, 191 ff.
 - extract of, 196
 - glucose fermentation by, 276
- Zero, absolute, 45
- Zinc, 124
 - electrons and, 127, 128
 - enzymes and, 257
- Zinc iron, 126
- Zinc sulfate, 127
- Zymase, 255